

# UNIVERSITY OF NAPLES “FEDERICO II”

## PhD Program

### Molecular Pathology and Physiopathology



School of Molecular Medicine  
XXVII Cycle

*Chromatin modification induced by DNA damage and repair.  
DNA methylation induced by DNA damage and repair causes  
stochastic variations of gene expression.*

SUPERVISOR  
Prof. Vittorio Enrico Avvedimento

CANDIDATE  
Dr. Giusi Russo

PhD COORDINATOR  
Prof. Vittorio Enrico Avvedimento

Academic Year 2013-2014

*Non è la specie più forte a sopravvivere e nemmeno la più intelligente.  
Sopravvive la specie più predisposta al cambiamento.*

*Charles Darwin*

***For you, who are and will be the my reason for living ...***

*Chromatin modification induced  
by DNA damage and repair.  
DNA methylation induced by DNA damage  
and repair causes stochastic variations  
of gene expression.*

# TABLE OF CONTENTS

	<b>Page</b>
<b>LIST OF ABBREVIATIONS.....</b>	<b>4</b>
<b>ABSTRACT.....</b>	<b>5</b>
 <b>1. BACKGROUND.....</b>	 <b>6</b>
1.1. DNA Damage: Double strand breaks.....	6
1.2.Repair of Double-Strand Breaks.....	8
1.3.Repair of double strand break by Gene Conversion.....	9
1.4.DNA Methylation .....	10
1.5.DNA damage and DNA methylation.....	11
 <b>2. AIM OF MY STUDY.....</b>	 <b>13</b>
 <b>3. MATERIAL AND METHODS.....</b>	 <b>15</b>
• Cell culture.....	15
• Vectors.....	16
• RNA and DNA extraction.....	17
• FACS analysis.....	18
• Bisulfite DNA preparation, PCR and sequence analysis.....	18
• Chromatin Immuno-Precipitation (ChIP).....	19
• Methylated DNA immunoprecipitation (MeDIP).....	20



• Chromosome conformation capture (3C).....	20
• Statistical analysis.....	22
• Buffers Formulation.....	22
<b>4. RESULTS.....</b>	<b>24</b>
4.1. Transient and permanent chromatin modifications induced by DNA damage and repair.....	24
• Histone code change induced by damage.....	24
• Histone code change induced by repair.....	25
• DNA Methylation stabilizes H3K9m2/3.....	26
4.2. Chromatin looping induced by damage and repair.....	27
4.3. BER enzymes remodel DNA methylation during HR.....	33
4.4. Homologous targeting of GFP in ES cells generates clones with various levels of GFP expression and DNA methylation.....	41
4.5. Discrete DNA methylation patterns mark clones with distinct gene expression levels.....	44
<b>5. DISCUSSION.....</b>	<b>48</b>
5.1. Transient and stable chromatin changes induced by DNA damage and repair.....	48
5.2. BER enzymes remodel DNA methylation during HR.....	50
5.3. Relation between methylation profiles and levels of gene expression.....	51

<b>6. CONCLUSION.....</b>	<b>52</b>
<b>7. ACKNOWLEDGMENT .....</b>	<b>53</b>
<b>8. REFERENCES .....</b>	<b>54</b>
<b>LIST OF PUBLICATIONS.....</b>	<b>59</b>

## LIST OF ABBREVIATIONS

**3C** Chromosome conformation capture  
**APE1** Apurinic/apyrimidinic endonuclease  
**ATG** Translation Start Site  
**BER** Base-Excision Repair  
**ChIP** Chromatin Immuno-Precipitation  
**Dam** DNA adenine methylase  
**DDR** DNA-damage response  
**Dnmt1** DNA (cytosine-5) methyltransferase-1  
**DNMTs** DNA-methyltransferase  
**DSB** Double-Strand Break  
**HR** Homologous Recombination  
**MeDIP** Methylated DNA immunoprecipitation  
**MMR** Mismatch Repair  
**NER** Nucleotide Excision Repair  
**NHEJ** Non-Homologous End-Joining  
**OGG1** 8-oxoguanine glycosylase,  
**SDSA** Synthesis-Dependent Strand Annealing  
**ssDNA** single-stranded DNA  
**TDG** Thymine-DNA glycosylase

## ABSTRACT

We report the local changes in chromatin structure and gene expression after homology-directed repair (HR) of a DNA double-strand break (DSB). The histone code at the DSB site undergoes transient changes following DNA damage, and is stably modified after repair. Chromatin changes are associated with modification in DNA methylation and local chromatin structure, which determine the rate of transcription of the repaired gene. During repair discrete DNA loops form that connect the 5' and the 3' ends of the repaired GFP gene. Some of these loops are sensitive to RNaseH, suggesting that they are stabilized by RNA:DNA hybrids. BER-NER enzymes recruited to the DSB influence the final methylation status of the repaired gene, as does transcription of the gene during the three weeks following repair. We also report that methylation-induced by homologous repair occurs also, during homologous targeting of genes in ES cells.

The changes in methylation patterns are permanent, and can be used to decode the cell history of past damage-repair events. If the expression of the methylated gene is under strong negative selection, the frequency of methylated alleles (epi-alleles) increases through cell generations. We analyzed the various epi-alleles of the GFP gene in terms of location of methylated CpGs and found that there is a tight association between the location of CpG methylated and the levels of expression of the gene.

These data demonstrate that somatic DNA methylation induced by DNA damage and HR stably modifies local chromatin structure and is a source of permanent stochastic variation of gene expression in somatic cells.

# **1. INTRODUCTION**

## **1.1. DNA Damage: Double strand breaks**

Human cells are targets of an enormous number of DNA lesions per day (Lindahl T et al.2000). These lesions inhibit DNA replication and transcription, and if they are not properly repaired they can generate genome aberrations that affect organism viability. Some DNA lesions are induced by endogenous physiological processes, such as DNA base alterations randomly introduced during DNA replication or transcription. DNA damage is also produced by reactive oxygen species deriving from oxidative respiration or by environmental toxins and heavy metals (Valko M. et al. 2006). Reactive oxygen and species are also produced by inflammation and infections because innate-immunity cells undergo oxidative bursts to kill bacteria (Kawanishi S. et al. 2006). Furthermore, collisions between RNA and DNA polymerases may generate double strand breaks, which are extremely dangerous and toxic for the cell.

To repair DNA damage, cells have evolved mechanisms – collectively termed the DNA-damage response (DDR) – to detect DNA lesions, signal their presence and promote their repair (Harper J.W. et al. 2007). Cells defective in these mechanisms generally display heightened sensitivity towards DNA-damaging agents and many such defects cause human diseases. The wide diversity of DNA-lesion types necessitates multiple, largely distinct DNA-repair mechanisms. While some lesions are subject to direct protein-mediated reversal, most are repaired by a sequence of catalytic events mediated by multiple proteins. In mismatch repair (MMR), detection of mismatches and insertion/deletion loops triggers a single-strand incision that is processed by nuclease, polymerase and ligase enzymes (Jiricny J. 2006).

The excision systems repair are: the base-excision repair (BER), where a damaged base and the nucleotide excision repair (NER) system, which recognizes lesions that distort the helix, operates via two sub-pathways that differ in the mechanism of lesion recognition: transcription-coupled NER, which specifically targets lesions that block transcription, and global-genome NER. A peculiarity of NER system is that the damage is excised as a base oligonucleotide, producing single-stranded DNA (ssDNA) that is processed by DNA polymerases and associated factors before ligation (Hoeijmakers J.H.J. 2001). Also, some DNA lesions are not repaired but are instead bypassed during DNA replication by low fidelity polymerases with less stringent base-pairing requirements than replicative polymerases (Loeb L.A. et al., 2008).

## **1.2. Repair of Double-Strand Breaks**

The most dangerous form of DNA damage for the cell is the Double - Strand Break (DSB). The DSBs are lesions that can cause problems for DNA transcription, replication and chromosome segregation and may result in genomic instability. Genomic instability is the major cause of cancer through the activation of proto-oncogenes or inactivation of tumor suppressor genes (Essers J. et al 2000).

The DSB are generated when the two complementary strands of the DNA double helix are broken simultaneously in neighboring sites so that the coupling of the bases and the structure of chromatin become insufficient to maintain the ends of the DNA juxtaposed. Dissociated ends can recombine with other genomic sites leading to mutations and chromosomal aberrations.

Eukaryotic cells have complex systems that detect and signal the presence of these lesions.

The two main systems deputies to the repair of DSB are:

1. Repair mediated homologous recombination (Homologous Recombination or HR);
2. Repair of non-homologous ends (Non-Homologous End Joining or NHEJ).

These mechanisms, although very different, act in a strictly complementary fashion and are significantly conserved during evolution (EssersJ.et al. 2000; Cromie et al. 2001).

The choice of which system to use seems to be influenced by the stage of the cell cycle in which the cell when the damage occurs. In eukaryotes, homologous recombination is restricted to S phase or G2 phase of the cell cycle, when the double-strand breaks are repaired using the long regions of homology (> 100bp) available from sister chromatids during separation. Non-homologous recombination appears to be the dominant mechanism in multicellular eukaryotes, active during all phases of the cycle (Burma et al. 2006).

During the HR, the damaged chromosome form a pairing complex with a DNA molecule not damaged with which it shares the sequence homology. Under these conditions the genetic information lost on one allele can be recovered on the other, remained intact. The enzymatic machinery responsible for this process is identical to the one in charge for the crossing-over during meiosis. The steps of the repair by homologous recombination are: 1. processing of the ends to single strand; 2. homology pairing; 3. copying the information from the non-damaged strand; 4. resolution of the complex and reconstitution of the double helix.

The NHEJ, instead, connects the two ends of a DSB through a process that is independent of the terminal sequence homologous DNA and therefore produces joints which can vary in their sequence.

It's unknown the precise factors that determine which system will repair the DSB. If a DSB occurs during replication, the local forks arrest and the DNA damage checkpoint is activated. HR seems the preferred mechanism to repair

DSB during replication. Inefficient HR under these conditions may increase the replication stress (McCabe et al. 2006).

### **1.3. Repair of double strand break by Gene Conversion**

Gene conversion is an event of homologous recombination of DNA that occurs mainly during the meiotic division, but can also occur in somatic cells and does not involve cross-over. It is a non-reciprocal recombination mechanism which keeps the genetic identity and also promotes diversity (Santoyo et al. 2005).

During gene conversion, the information contained in the sequence of DNA is transferred from a strand which acts as donor and which remains unchanged to the other strand with which the first share a certain sequence homology. The DNA sequences involved are homologous but not identical: the result will be the generation of two DNA sequences identical to that specific locus that has suffered damage.

The most frequent mechanism of gene conversion is the Synthesis-Dependent Strand Annealing (SDSA), in which there is no cross-over and the single strand is annealed and copied by the complementary strand. Under these conditions there is transfer of information but not physical exchange.

Gene conversion serves to make uniform the DNA sequences that comprise the genes of one species and, with the passage of time, they create homogeneous groups of DNA sequences, a phenomenon that affects both allelic forms of a gene, and multigene families.



## 1.4. DNA Methylation

DNA methylation is a post-replicative covalent modification of genomic DNA consisting in the introduction of a methyl group (-CH<sub>3</sub>) on carbon 5 of the nitrogenous base cytosine giving rise to 5-methyl-cytosine. Methylation has an important role in the regulation of gene expression in eukaryotes (Eden and Cedar 1994). It is essential for the development and differentiation of mammals (Li et al. 1992), for the inactivation of the X chromosome (Panning and Jeanisch 1998) and is involved in the imprinting (Li et al. 1993) and in the tumorigenesis (Laird et al. 1995). Methylation facilitates the distinction of the newly synthesized DNA strand from parental and, at the same time, compacts and protects the cell genome. While in bacteria the adenine can be methylated only at the level of the genome sequences GmeATC Dam enzyme (DNA adenine methylase); in mammals, methylation occurs only on carbon C5 cytosine belonging to CpG dinucleotide (cytosine-guanine), by specific enzymes called DNA methyltransferase. There are clusters of CpG in the genome called CpG islands.

Generally, in mammals CpG islands are located within and close to the promoters in 40% of genes (it gets to 70% in human promoters). In vertebrates CpG islands are located near or at the site of initiation of transcription of the genes, in particular those defined housekeeping. CpG islands bearing aberrant mutations were observed in various tumors, where they are often hyper-methylated (Baylin et al. 2000; Yan et al. 2003). Typically cells and cancerous tissues show a wide hypomethylation of the genome and a localized hyper-methylation. DNA methylation contributes to the formation of inactive chromatin and gene silencing (Bird and Wolff 1999). The methylation of DNA in mammalian cells is regulated by a family of DNA-methyltransferase (DNMTs): DNMT1; DNMT3a; DNMT3b.

DNMT1 copies the methylation profile of the parent strand of the newly synthesized after each replication cycle, and is responsible for the transmission

and maintenance of methylation patterns established during embryonic development and cell division. DNMT3a and DNMT3b instead have the ability to add methyl groups to DNA that has no modification and are therefore referred to as "de novo methyltransferases". They are expressed ubiquitously in proliferating cells. The inactivation of DNMT1 is lethal and in embryonic mouse stem cells induces massive demethylation (Li et al. 1992), without interfering with the proliferation (Hong et al. 1996).

### **1.5. DNA damage and DNA methylation**

In the laboratory it was demonstrated by using a system pioneered by M. Jasin (Jasin M. 1996; Pierce A.J. et al. 1999), in which recombination between partial duplications of a chromosomal GFP gene is initiated by a specific DSB in one copy. The unique DSB is generated by cleavage with the meganuclease I-SceI, which does not cleave the eukaryotic genome. The DSB is repeatedly formed and repaired, until the *I-SceI* site is lost by homologous or non homologous repair or depletion of I-SceI enzyme. Recombination products can be detected by direct analysis of the DNA flanking the DSB or by the appearance of functional GFP (Cuozzo C. 2007).

Two cell types are generated after recombination: clones expressing high levels of GFP and clones expressing low levels of GFP, referred to as H and L clones, respectively. Relative to the parental gene, the repaired GFP is hypomethylated in H clones and hypermethylated in L clones. The altered methylation pattern is largely restricted to a segment just 3' to the DSB. Hypermethylation of this tract significantly reduces transcription, although it is 2,000 bp distant from the strong CMV promoter that drives GFP expression (Cuozzo C. 2007; Rountree M.R. 1997). The ratio between L and H clones is approximately 1 to 2 or 1 to 4, depending on the insertion site of the GFP reporter. These experiments were performed in mouse embryonic (ES) or

human cancer (HeLa) cells. HDR-induced methylation was dependent on DNA methyl transferase I (DNMT1). Furthermore, methylation induced by HDR was independent of the methylation status of the converting template (Cuozzo C. 2007). These data, taken together, argue for a cause-effect relationship between DNA damage-repair and DNA methylation.

The link between DNA damage, repair and *de novo* methylation has been confirmed by other studies (O'Hagan H.M. 2008; Ha K. et al.2011).Moreover, it is worth noting that genome wide surveys show that imprinted sites are historical recombination hot spots, reinforcing our conclusion and that of other workers, that DNA methylation marks the site of DNA recombination(Sandovici I. 2006; Sigurdsson M.I. 2009).

On the basis of these observations, I have decided to further explore the links between DNA damage-repair and DNA methylation with specific emphasis on the changes of chromatin during and after repair at the DSB and on the mechanism leading to methylation of the repaired DNA segment.

## 2. AIM OF THE STUDY

DNA methylation is not sequence-specific and it is not known how and when a specific segment of DNA becomes methylated by DNA methyl transferase enzymes. It is known that DNMT1 during replication methylates hemimethylated DNA at high efficiency and this mechanism explains the inheritance and stability of methylation profiles. However, the DNMT enzymes able to de novo methylate a segment of DNA are well known (DNMT3a-3b) but the mechanism leading to de novo methylation of DNA is still elusive. Our data indicate that DNA damage and homologous repair are, indeed the events leading to de novo methylation but it is not known how exactly it happens and which are the chromatin modifications underlying DNA methylation during repair.

By using the DRGFP system illustrated above, we wish to approach the following questions:

1. Chromatin modifications during and after repair and their association with de novo methylation of the damaged DNA segment. This will be achieved by the analysis of the changes of histone code in chromatin where the DSB occurs, before, during and after the homologous repair. The association of the modifications of chromatin and de novo methylation will be analyzed by perturbing repair and/or DNA methylation.

2. Is DNA methylation after repair responsible for variation of gene expression? Cells after repair bear identical genotypes but different expression levels of the repaired gene. I wish to compare the methylation profile and the expression levels in isolated clones to find a possible relation.

3. Polymorphism of methylation in cell populations may serve as guide to identify and track specific epialleles, i.e. alleles differing only by methylation in specific DNA segments. This analysis can be performed in genes that are silenced in cancer.

### 3. MATERIALS AND METHODS

- **Cell culture**

HeLa cells lines were cultured at 37 °C in 5% CO<sub>2</sub> in RPMI medium supplemented with 10% fetal bovine serum (Invitrogen), 1% penicillin-streptomycin, and 2 mM glutamine. HeLa-DR-GFP cells were obtained by stable transfection of HeLa cells with the pDR-GFP plasmid. HeLa-DR-GFP stable lines were transfected with lipofectamine as recommended by the manufacturer (Invitrogen) with 2 µg of circular pDR-GFP plasmid and selected in the presence of puromycin (2 mg/ml). Puromycin-resistant colonies (approximately 200 clones) were seeded at 3 x 10<sup>5</sup> cells per 60-mm plate and transfected with 2.5 µg pCbASce plasmid DNA on the following day by lipofectamine transfection(Invitrogen). Cells were harvested 2 or 4 d after transfection. Transfected cells were cultured for 2 weeks and analyzed. Pools of clones were generated in two or three independent transfections and frozen in aliquots. Transient transfections with I-SceI were carried at different times of culture after the primary transfection (for the isolation of single clones, we used cells cultured for more than 1 month). We used the same conditions of growth (~40%confluency starting from freshly frozen aliquots). Transfection efficiency was measured by assaying β-galactosidase activity of an included pSVbGal vector (Promega). Normalization by FACS was performed using antibodies to β-gal or pCMV-DsRed-Express (Clontech). pEGFP (Clontech) was used as GFP control vector.

Transient transfections with I-SceI were carried at different times of culture after the primary transfection (for the isolation of single clones, we used cells cultured for more than 1 month). We used the same conditions of growth (~40% confluency starting from freshly frozen aliquots). The structure of the pDR-GFP and other plasmids are described below.

- **Vectors**

pDR-GFP plasmid was 14,735 bp containing (see Morano A. et al. 2014): CMV IE enhancer (1–385); chicken beta actin promoter (386-751); chicken beta-actin first intron (752-1,622); rabbit beta-globin second intron (1,623-1,670); rabbit beta-globin third exon (1,671-1,724); EGFP with a STOP (Pierce et al., 1999) codon at I-SceI site (1,740-2,756, I-SceI at 2,135); SV40 splice/polyadenylation signal (2,757-3,023); polyadenylation signal from phosphoglycerate kinase gene (3,025-3,607) for the puromycin resistance gene (3,600-4,200); a truncated EGFP gene sequence (5,609-6,138); and 6,450 bp of mouse genome (Jasin, 1996 and A. Porcellini, unpublished data).

siRNA against TDG: cat.# SR304768; OriGene Technologies, Inc., USA shRNA of APE1 and APE1WT plasmids. The following oligonucleotides were used for the short hairpin RNA (shRNA) of APE1:sense, 5'-TCCCCCTGCCACACTCAAGATCTGCTTCAAGAGAGCAGATCTTGA GAGTGGCAGGTTTTTGGAAA-3'; and antisense, 5'-GCTTTTCCAAAACC TGCCACACTCAAGATCTGCTCTCTTGAAGCAGATCTTGAGTGTGGC AGGGGG-3'. These sequences were drawn following the empirical rules of Mittal and were designed to recognize and bind to a 21-base sequence (underlined) placed 175 nucleotides after the AUG initiation codon of the APE1 gene. As a control, we used the following scrambled oligonucleotide sequences: sense, 5'-CGCCACCCCGTATTCAAGAGATACGGGTGGCGAG TTAGACTTTTTTGGAAA-3'; antisense, 5'-AGCTAAAAAAGTCTAACT CGCCACCCCGTATCTTGAATACGGGGTGGCGAGTTAGACTGGG-3'.

These sequences were checked with BLAST(<http://www.ncbi.nlm.nih.gov/blast/>) for their inability to pair with any human cDNA sequence. The sequences were cloned into Bgl II and Hind III restriction sites of pSUPER vector (Oligoengine) to form the so-called pSUPER-APE1 vector.

The pFLAG-CMV-5.1/APE1-M2 expression vector was generated by modifying the pFLAG-CMV-5.1/APE1 (Vascotto et al. 2009). To avoid the degradation of the ectopic APE1 mRNA by the specific siRNA sequence described above, two nucleotides of the APE1-cDNA coding sequence were mutated with site-directed mutagenesis kits (Stratagene), leaving the APE1 amino acid sequence unaffected: siRNA, 5'-CCTGCCACACTCAAGATCTGC-3'; and APE1, 5'-CCTGCAACGCTCAAGATCTGC-3'.

The expression vectors for OGG1WT and for the K338R/K341R OGG1mutant were the FLAG-Tagged vectors previously described by Bhakat (Bhakat et al, 2006).

### • RNA and DNA extraction

Total RNA was extracted using Triazol (Gibco/Invitrogen). Genomic DNA extraction was performed with following protocol: cellular pellet was resuspended in 10 mM TRIS (pH 7.8) and 50 mM NaCl solution (2x10<sup>7</sup> cells/ml). After addition of 1% SDS the sample was gently mixed. Proteinase K, at a final concentration of 200 ug/ml, was added and the mixture was incubated at 55 °C overnight. The following day, hot NaCl solution (70 °C) was added to the mixture at final concentration of 1.5 M and the DNA was extracted by chloroform. DNA was ethanol precipitated, dried, and resuspended in TE buffer. Total RNA was extracted using TRIzol (Gibco/Invitrogen). cDNA was synthesized in a 20 ul reaction containing 1 ug of total RNA, 100 U of Superscript III Reverse Transcriptase (Invitrogen), and 2 ul random hexamer (20 ng/ul) (Invitrogen). mRNA was reverse-transcribed for 1 h at 50 °C, and the reaction was heat inactivated for 15 min at 70 °C. The products were stored at -20 °C.



- **FACS analysis**

For the FACS analysis, HeLa-DR-GFP cells were harvested and resuspended in 500 µl of PBS at density of 10<sup>6</sup> cells/ml. Cell viability was assessed by Propidium Iodide (PI) staining: before FACS analysis cells were incubated with 3 µM PI for 10 min. Cytofluorimetric analysis was performed on a 9600 Cyan System (DakoCytometrix), PI positive cells were excluded from the analysis by gating the PI-negative cells on a FSC-Linear vs FL2H-Log plot; GFP<sup>+</sup> cells were identified with a gate (R1) on a FL1H-Log vs SS-Log plot. Rec L and Rec H cells were identified on a FL1H Histogram of the R1-gated cells with 2 range-gate (see Fig. 5A). The same gate was used for all the cytofluorimetric analysis. Population comparison was performed using the FlowJo software (Chi-Squared Test). Differences in fluorescence intensity (mean) were determined using the matched pairs Student's t test.

- **Bisulfite DNA preparation, PCR, and sequence analysis**

Sodium bisulfite analysis was carried out on purified genomic DNA as described. Chromatinized DNA was obtained as follows: 10<sup>7</sup> cells were fixed at 4°C temperature with 1% formaldehyde for 3 min. The reaction was stopped with glycine to a final concentration of 125 mM. Nuclei were isolated and permeabilized by incubating cells for 20 min in Buffer A, 20 min in Buffer B and then resuspended in Buffer C (see Buffers Formulation). Nuclei or purified genomic DNA was heat denatured (96 °C for 10 min) incubated in a fresh solution containing 5 M sodium bisulfite and 20 mM hydroquinone and incubated at 37°C for 18 h. The cross-link was reversed, and proteins were digested with proteinase K (50 µg/ml at 55°C for 2 h, and then at 65°C overnight). DNA was purified using a Wizard genomic purification kit (Promega), and then disulphonated by incubation for 15 min with NaOH to a

final concentration of 0.3 M, neutralized with ammonium acetate to a final concentration of 3 M, and purified by ethanol precipitation.

DNA was amplified by PCR using primers listed in the Table T1 using Taq polymerase, which is able to copy deoxyuridine, cloned in TOPO TA vectors (Invitrogen), and sequenced with the M13 reverse primer. Sequence analysis and alignment were performed using MegAlign software (a module of the Lasergene Software Suite for sequence analysis by DNASTAR) for MacOSX. Statistical analysis was performed using the JMP 6.0.3 statistical analysis software by S.A.S.

- **Chromatin Immuno-Precipitation (ChIP)**

Cells were transfected and/or treated as indicated in the legends of the figures. The cells ( $\sim 2.5 \times 10^6$  for each antibody) were fixed for 10 min at room temperature by adding 1 volume of 2% formaldehyde to a final concentration of 1%; the reaction was quenched by addition of glycine to a final concentration of 125 mM. Fixed cells were harvested and the pellet was resuspended in 1 ml of LysisBuffer containing 1X protease inhibitor cocktail (Roche Applied Science). The lysates were sonicated to have DNA fragments 300 to 600bp. Sonicated samples were centrifuged and supernatants diluted 2 fold in the ChIP Buffer. An aliquot (1/10) of sheared chromatin was further treated with proteinase K, extracted with phenol/chloroform and precipitated to determine DNA concentration and shearing efficiency (input DNA). The ChIP reaction was set up according to the manufacturer's instructions. Briefly, the sheared chromatin was precleared for 2 h with 1 ug of non-immune IgG and 20 ul of Protein A/G PLUS-Agarose (Santa Cruz Biotechnology) saturated with salmon sperm (1 mg/ml). Precleared chromatin was divided in aliquots and incubated at 4 °C for 16 h with 1 ug of the specific antibody and non-immune IgG respectively. The immuno-complexes were recovered by incubation for 3

h at 4°C with 20 µl of protein-A/G agarose, beads were washed with wash buffers according to the manufacturer's instructions and immunoprecipitated DNA was recovered and subjected to qPCR using the primers indicated in the legend of the specific figures, primers sequences are described in Table T1.

- **Methylated DNA ImmunoPrecipitation (MeDIP)**

Cells were transfected and/or treated as indicated in figure legends. A total of  $\sim 5 \times 10^6$  cells were harvested and Genomic DNA extracted as described above. Ten micrograms of total genomic DNA was digested in 200 µl for 16 h with Restriction Endonuclease mix containing 30 U each of Eco RI, Bam HI, Hind III, XbaI, Sal I (Roche Applied Science), phenol/chloroform extracted, ethanol precipitated and resuspended in 50 µl of TE buffer. An aliquot (1/10) of digested DNA was used as input control to determine DNA concentration and digestion efficiency. MeDIP was performed essentially as described (Lee B et al. 2011) except typically 2 µg of antibody specific for 5mC (Abcam) was used to precipitate methylated DNA from 5 µg of total genomic DNA.

- **Chromosome conformation capture (3C)**

The actual position of oligonucleotides and Hae III restriction sites are shown in Fig. 2A.

A total of  $5 \times 10^6$  cells were crosslinked in 20 ml of serum-free medium with 1% formaldehyde for 10 min at room temperature. The reaction was quenched by the addition of glycine to a final concentration of 125 mM. Fixed cells were harvested and the pellet resuspended in 0.5 ml of ice cold lysis buffer (10 mM Tris pH 8.0, 10 mM NaCl, 0.2% NP40 and protease inhibitors), and incubated on ice for 15 min. Nuclei were washed with 0.25 ml of restriction enzyme

buffer and pelleted. Nuclei were then resuspended in 180  $\mu$ l of restriction enzyme buffer. SDS was added to a final concentration of 0.1%, and nuclei were incubated at 37 °C for 15 min. Triton X-100 was then added to the final concentration of 1% to sequester SDS. Digestion was performed with 150 U of Hae III restriction enzyme at 37 °C for 16 h. The restriction enzyme was inactivated by the addition of SDS to 2% and incubation at 65 °C for 30 min. The reaction was diluted into 4 ml ligation reaction buffer containing 50 U of T4 DNA Ligase (Roche Applied Science). Ligations were incubated at 16 °C for 18 h. EDTA (to a concentration of 10 mM) was added to stop the reactions. Samples were treated with Proteinase K (200  $\mu$ g/ml) and incubated for 2 h at 50 °C, and then overnight at 65 °C to reverse the formaldehyde crosslinks. The following day, the DNA was purified by phenol/chloroform extraction and ethanol precipitation. Samples were redissolved in 20  $\mu$ l of TE buffer. To prepare a control template, we used pDR-GFP plasmid. Five micrograms of plasmid DNA were digested with HaeIII in 50  $\mu$ l of 1x buffer for 8 h at 37 °C and then ligated in 20  $\mu$ l with 5 U of T4 Ligase at 16 °C for 4 h. The efficiency of digestion after the entire 3C treatment was quantified by Real Time PCR, amplifying a fragment spanning two Hae III sites (uncut) in different 3C DNA preparations. A total of 35 rounds of PCR amplification were used. The efficiency of ligation was assayed as follow: a linear Hind III-digested pUC18 plasmid was added to all the preparations before ligation. The ligated plasmid was quantified by real-time PCR, amplifying a fragment spanning the Hind III site (M13/pUC Forward and Reverse primers were used). An appropriate amount of DNA that would amplify within the linear range was subsequently used for the experiments.

The efficiency of digestion was quantified by Real Time PCR, amplifying three fragments spanning three different Hae III sites. The presence of the amplified products indicated the efficiency of digestion, and was expressed as ratio between cleaved/un-cleaved band. Experiments with a cutting Ratio >0.03

( $\Delta CT \geq 4$ , 97% digestion) were discarded. Primer sequences are available in Table T1.

- **Statistical analysis**

All data are presented as mean  $\pm$  standard deviation in at least three experiments in triplicate ( $n \geq 9$ ). Statistical significance between groups was determined using Student's *t* test (matched pairs test or unmatched test were used as indicated in figure legends). All tests were performed using the *JMP Statistical Discovery*<sup>TM</sup> software by SAS.

- **Buffers Formulation**

Bisulfite Buffer A: 10 mM Tris-HCl, pH 8, 10 mM EDTA, pH 8, 0.5 mM EGTA, pH 8, 0.25% Triton X-100. Bisulfite

Buffer B: 0.2 M NaCl, 10 mM Tris-HCl, pH 8, 1 mM EDTA, 0.5 mM EGTA.

Bisulfite Buffer C: 0.3 M NaCl, 40 mM Tris-HCl, 4 mM EDTA, 1% Triton X-100.

3C Buffer 1: 10 mM Tris pH 8.0, 10 mM NaCl, 0.2% NP40 and protease inhibitors.

3C Ligation Buffer: 66 mM Tris-HCl, 5 mM MgCl<sub>2</sub>, 5 mM DTT, 1 mM ATP, pH 7.5.

ChIP Lysis Buffer: 10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2 % NP40.

ChIP Buffer: 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.0.

ID	PRIMERS	Locus
3C Primer a	5'-GGCGGGGCGAGGCGGAGA-3'	pDR-GFP
3C Primer b	5'-GAGTCGCTGCGTTGCCTTC-3'	pDR-GFP
3C/ChIP Primer c	5'-CGCCCGCAGCGCTCACAGC-3'	pDR-GFP
3C/ChIP Primer d	5'-ACGTGCTGGTTATTGTGCTGTC-3'	pDR-GFP
3C/ChIP Primer e	5'-TCCTGCTCCTGGGCTTCTCG-3'	pDR-GFP
3C Primer f	5'-AAGATGGTGCGCTCCTGGACGTA-3'	pDR-GFP
3C Primer g	5'-GGTGAAGTTCGAGGGCGACAC-3'	pDR-GFP
3C Primer h	5'-TGCACGCTGCCGTCCTCG-3'	pDR-GFP
Primer Rec2	5'-CGGCGGCGGTACGAAGTC-3'	pDR-GFP
3C Primer i	5'-GCTGATCTCGTTCTTCAGGC-3'	pDR-GFP
3C Primer l	5'-GGTACTCTGTTCTCACCTTC-3'	pDR-GFP
3C Primer m	5'-GAAAGCGAAGGAGCAAAGCTG-3'	pDR-GFP
ChIP Primer a	5'-CGTTACTCCCACAGGTGAGC-3'	pDR-GFP
ChIP Primer r	5'-CCACCGGCAAGCTGCCCCGTGCC-3'	pDR-GFP
ChIP Primer h	5'-CTCGGCGCGGGTCTTGTAGTTGC-3'	pDR-GFP
ChIP Primer p	5'-GGCGTAAATTGTAAGCGTTA-3'	pDR-GFP
ChIP Primer q	5'-GCCTGAAGAACGAGATCAGC-3'	pDR-GFP
Primer I-SceI	5'-GCTAGGGATAACAGGGTAAT-3'	pDR-GFP
Primer Bcg	5'-GAGGGCGAGGGCGATGCC-3'	pDR-GFP
Bisulfite E01F	5'-GTGTGATTGGTGGTTTTAGAGT-3'	pDR-GFP
Bisulfite E02R	5'-CCATCCTCAATATTATAACAAAT-3'	pDR-GFP
Bisulfite E2F	5'-GGAGTTGTTTTATTGGGGTGGTGTATTTTTGGT-3'	pDR-GFP
Bisulfite E2NF	5'-TGGATGGTGATGTAAATGGTTATAAGTTT-3'	pDR-GFP
Bisulfite E2R	5'-GTTTGTGTTTTAGGATGTTGTTG-3'	pDR-GFP
Bisulfite E4R	5'-ACTTATACAACATCCATACCAAAAATAATCC-3'	pDR-GFP
Bisulfite E5R	5'-ACTTATACAACATCCATACCGAAAATAATCC-3'	pDR-GFP
Bisulfite E6R	5'-GGTTGTTATGAATAAAGGTGGTTATAAGA-3'	pDR-GFP
Bisulfite E7R	5'-CTCACTCATTAACACCCCCAACTTTACAC-3'	pDR-GFP
Bisulfite E8R	5'-GAAGATTTTTPyGATTTGTAGTTTAAGTTTATAGG-3'	pDR-GFP
Bisulfite E9R	5'-GAAGATTTTTPyGATTTGTAGTTTAAGTTTATAGG-3'	pDR-GFP
MEDIP I F	5'-TGTTTTGGTTGGCGTAAGGC-3'	Mouse Rosa26
MEDIP I R	5'-CGGCGGGTAAACGACTCC-3'	Mouse Rosa26
MEDIP II F	5'-AGAAAAGTGGCCCTTGCCAT-3'	Mouse Rosa26
MEDIP II R	5'-TTGAGGCAACTCAAGTCGGAA-3'	Mouse Rosa26
18SF	5'-TCCCCATGAACGAGGAATTC-3'	h18S
18SR	5'-GTGTACAAAGGGCAGGGACTT-3'	h18S

**Table T1: Complete list of DNA oligonucleotides used for PCR.** On the left is show in the primer identification tag and the specific experiments (ID); on the center, the DNA sequence; on the right, the genes or loci corresponding to the specific primer.

## 4. RESULTS

### 4.1. Transient and permanent chromatin modifications induced by DNA damage and repair

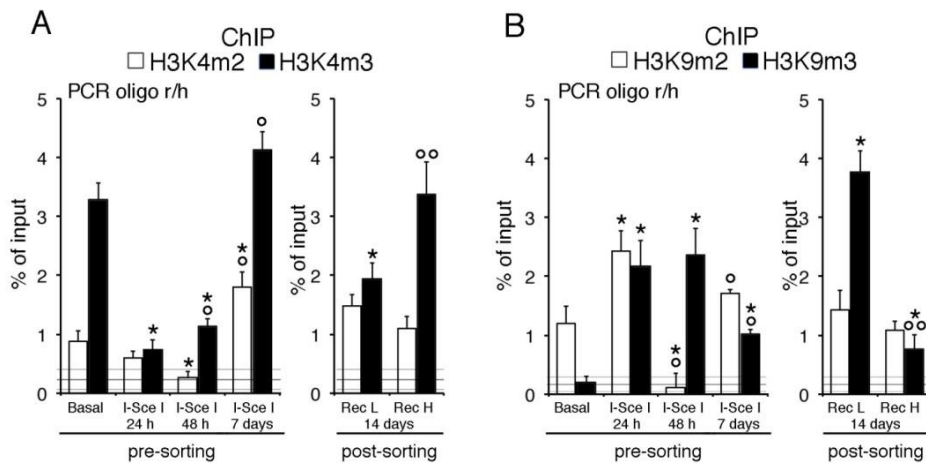
To determine the changes of chromatin proteins and DNA methylation induced by DNA damage and repair, we used cells, carrying stably integrated in the genome the specific construct of duplicated GFP, illustrated above, and transiently expressing the meganuclease I-SceI. We have initially characterized and determined the timing of damage (DSB) and repair (GFP fluorescent cells) following the expression of the nuclease. In fact, 72 h after expression of the gene coding for I-SceI, the repair is complete and the enzyme disappears from the mass cell culture. Taking into account this information, we have measured the modifications of histone H3 (histone code) after a short (24h) and a long time (48h) from I-SceI transfection.

- **Histone code change induced by damage**

Chromatin was precipitated with antibodies recognizing H3K4m2/3 and H3K9m2/3 and the immunoprecipitated DNA was amplified with primers corresponding to a fragment common to recombinant and non-recombinant GFP gene. Methylation of H3K4me2/3 at chromatin surrounding the DSB site was lost 24-48 h after exposure to I-SceI and returned to normal levels at 7 days later (Fig. 1A). Conversely, H3K9 methylation at the DSB site was significantly stimulated at 24 and 48 h after transfection with I-SceI, and decreased towards basal levels 7 days later (Fig. 1B). Since this modification occurs very early after DSB formation and before HR can be detected (24 h), we suggest that is dependent on DNA damage and not on repair.

- **Histone code change induced by repair**

To determine the long-term effects of HR on the H3K4m2/3 and H3K9m2/3, we sorted Rec H and Rec L cells, the cells that are generated after HR, corresponding to high (low methylation) and low (high methylation of the repaired gene) expressors of the repaired GFP gene. The methylation status of H3K4 and H3K9 at repaired GFP chromatin in Rec H cells was similar to that found before or 7 days after DSB formation (Fig. 1A and B). Rec L cells, on the other hand, retained H3K9m3 at the *I-SceI* site (Fig. 1B). Thus, H3K9m3 selectively marks the repaired gene in Rec L cells and represents a permanent, HR associated modification of the histone code. Recall that Rec L cells gained new methylated CpGs at the DNA region just 3' to the DSB after recombination. The data suggest that the DNA methylation and H3 methylation are functionally associated (Morano A. et al. 2014).



**Figure 1. DSB Histone code (H3K4m2/3 and H3K9m2/3) in cells undergoing DSB and repair.** A/B. HeLa DRGFP were transfected with I-SceI and analyzed 24, 48h and 7 days later. Cells were fixed and the chromatin analyzed for ChIP with the antibodies indicated. qPCR on each immunoprecipitate was carried out with r/h primers, which are common to recombinant and un-recombinant GFP DNA (see Figure 2). Pre-sorting and post-sorting indicate the chromatin derived from the mass culture at 24 and 48 h after I-SceI transfection or from cells (H and L) sorted after I-SceI transfection, respectively. Sorting was carried out 7 days after transfection and the cells were analyzed one week later. \*P<0.01 as compared with untreated control or basal. These data were normalized to total H3 histone. \*P<0.01 as compared with the previous time point; \*\*P<0.01 as compared versus Rec L.

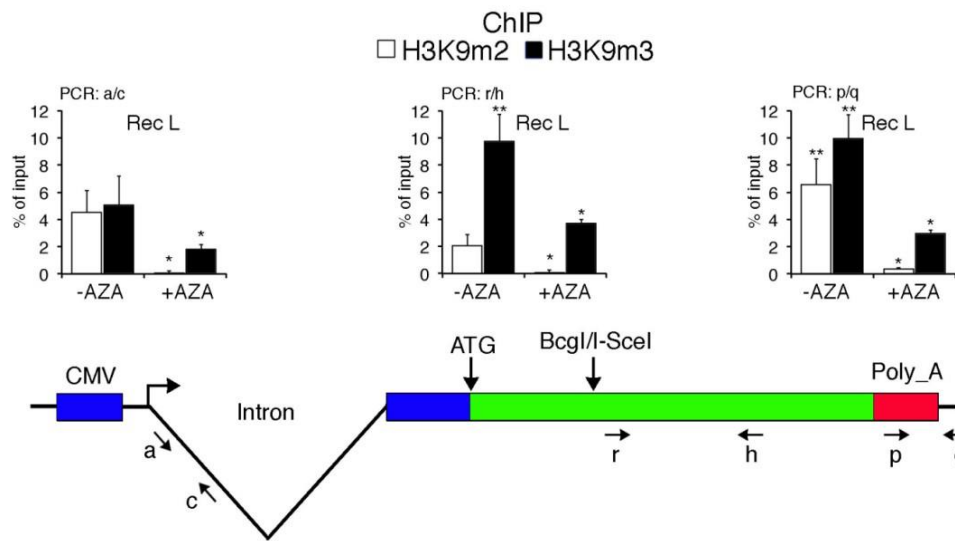


- **DNA Methylation stabilizes H3K9m2/3**

It has been reported that H3K9m3 facilitates recruitment of chromatin modifiers (for example, TIP60) that control DNA damage signaling (Sun Y. et al. 2009; Vaissiere T. et al. 2010) and suppress local transcription during repair (Stewart M.D. et al. 2005).

Why does the elevated H3K9m3 mark persist in Rec L cells? We hypothesize that HR-induced methylation of the repaired DNA may be responsible for H3K9m3 maintenance at the repaired site. To test this idea, we treated Rec L cells with 5-azadC and measured H3K9m2 and m3 at the three regions of the GFP gene indicated in Fig. 2. We find that 5-azadC significantly reduced H3K9m2 and m3 content not only at the repaired site, but also throughout the gene (Fig. 2*BcgI/I-SceI*). Since the promoter and the polyA regions contain few methylated CpGs, we suggest that the loss of methylation at the *BcgI/I-SceI* site influences local chromatin structure and H3K9 methylation.

Collectively, these data indicate that the rapid, massive and transient increase of H3K9m2/3 is induced by DSBs in all cells. Seven days after DSB formation, this mark disappears in the majority of cells (repaired by NHEJ or HR) and is replaced by H3K4m2/3. Only in Rec L cells, in which GFP DNA is methylated following HR, does H3K9m2/3 remain at the DSB site.

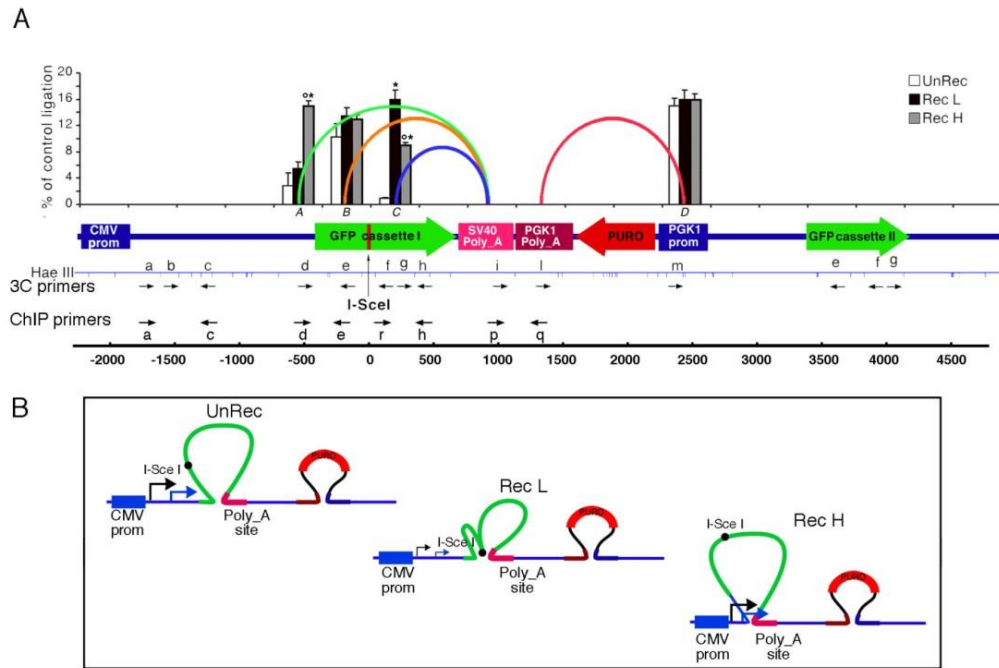


**Figure 2. DSB and HR stably modify the histone code at GFP locus.** Histone code (H3K9m2/3) of GFP in the chromatin derived from L or H sorted clones, untreated or treated with 5-azadC (10  $\mu$ M for 3 days and analyzed 4 days later). ChIP analysis was performed with the antibodies indicated. qPCR on each immunoprecipitate was carried out with primers r/h, indicated in the schematic diagram below the histograms. These data were normalized to total H3 histone. \*P<0.01 as compared with untreated control; \*\*P<0.01 as compared with the value found at promoter site (primers a/c).

## 4.2. Chromatin looping induced by damage and repair

DNA methylation at the site of DSB repair stabilizes the H3K9m3 mark, which may alter the local chromatin structure of the entire GFP gene (Fig. 2). To find specific chromatin domains induced or modified by damage and repair, we examined the structure of chromatin at the repaired locus in a pool of clones by chromosome conformation capture (3C). We consistently detected 4 major loops. One loop connected the PGK1\_polyA addition site of puromycin\_acetyltransferase gene with its promoter (loop D in red, Fig 3A); this loop is present in all cell types, marks the border of the puromycin\_acetyltransferase gene transcription unit, and represents an internal positive control in various experiments. The other loops have a common 3' end

(SV40/polyA addition site) and three different 5' ends, indicated as *A*, *B* and *C* (Fig. 3A). *A*, *B* and *C* mark different regions of the GFP "cassette I" at the 5' end, relative to the *I\_SceI* site, and are: 1. A region that includes a GFP transcription start site, identified by the primer *d* (loop *A*, green, from \_517 to \_279); 2. An internal region located upstream to the *I\_SceI* site, identified by the primer *e* (loop *B*, orange, from \_269 to \_30); and 3. A more distal 3' region of the GFP coding sequence located downstream of the *I\_SceI* site, including the boundary between the first and second GFP cassette, identified by primers *f* and *g* (loop *C*, blue, from +70 to +300). The frequency of loop *A* was high in Rec H and low in Rec L cells, whereas the abundance of loop *C* was low in Rec H and high in Rec L cells (Fig. 3A and B). Loop *A* brings the two transcription start sites (arrows in Fig. 3B) into close proximity to the polyA site. Loop *C*, instead, introduces a new DNA chromatin loop inside loop *B*, generating a novel structure involving a segment of GFP that is *de novo* methylated after repair (Fig. 3B). Since loop *C* is selectively abundant in Rec L cells, we hypothesize that the methylation status post HR may be responsible for the different 5' loop ends of Rec H (loop *A*) and Rec L (loop *C*) cells.



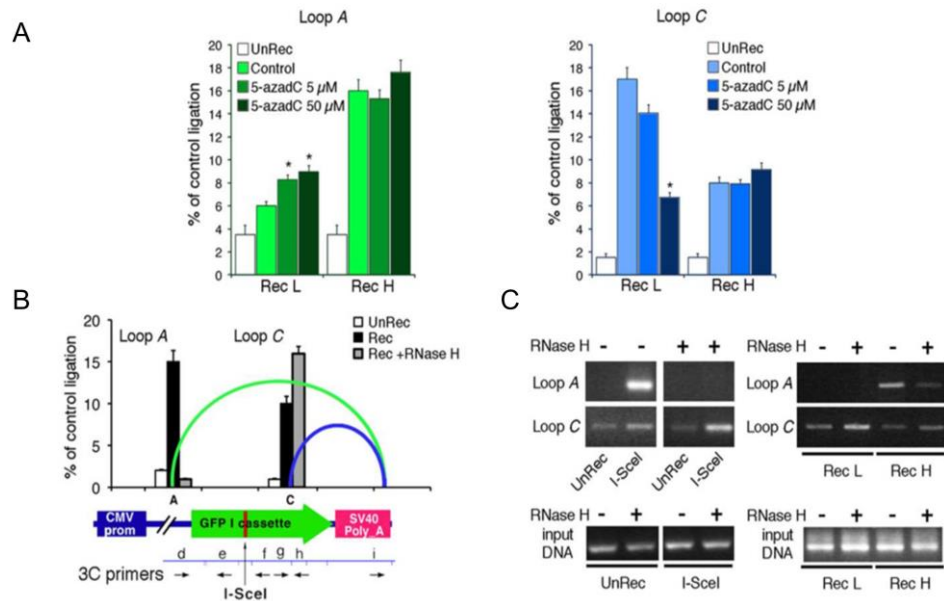
**Figure 3. 3C analysis of DRGFP chromatin in pool of clones subjected to DSB and HR.**

DNA was extracted from recombinant sorted H and L GFP positive cells (Rec H and Rec L) or control cells (UnRec) **A**. The histogram shows the frequency of ligation of the DRGFP HaeIII fragments amplified with primers indicated below the HaeIII restriction map. Ligation efficiency is relative to the DRGFP plasmid digested with HaeIII, ligated and amplified with specific primers (mean  $\pm$  SE). All the combinations of primers indicated, were performed on ligated chromatin; the histogram shows only loops with frequency above 1%, relative to the control. Distance, in bp, is relative to *I-SceI* site (vertical arrow). Each loop is identified by a capital letter; The loops A, B and C displays a unique (common) 3' end border (SV40/poly\_A addition site) and three different 5' ends: the 5' end of the loop A (green) includes a GFP transcription start site, identified by the primer d; loop B, (orange), includes a 5' region located upstream to the *I-SceI* site, identified by the primer e; the loop C, (blue), had a 5' end downstream of the *I-SceI* site including the boundary between the first and second GFP cassette, identified by primers f and g. The loop D (red) connects the PGK1\_poly\_A addition site with its promoter (PGK1 promoter). Differences between non-recombinant, H and L cells were tested by Student's *t* test: \* $p < 0.01$  as compared to UnRec; ° $p < 0.01$  as compared to Rec L cells. **B**. A cartoon that summarizes the formation of the loops in non recombinant and recombinant H and LGFP positive cells. Transcription start sites are indicated by arrows; the promoter, CDS and Poly\_A regions are indicated by a thick line using the same color code used in the upper panel.

To determine if loop C is dependent on hyper-methylation of the GFP region induced by repair, we treated Rec H and Rec L cells with 5-azadC for 72h and monitored formation of loops A and C. Fig. 4A shows that 5-azadC selectively inhibited the appearance of loop C in Rec L cells. In contrast, the

abundance of loops *C* and *A* in Rec H cells was resistant to DNA demethylation. Fig. 4A also shows that the relative abundance of the two loops changes in opposite directions in response to 5-azadC. In Rec L cells treated with increasing doses of 5-azadC, loop *A* progressively replaced loop *C*, suggesting that loss of loop *C* indicates the fraction of Rec L that convert to Rec H cells. These data suggest that loop *A* represents a positive transcription structure present in non-recombinant and Rec H clones, whereas loop *C* defines a structure associated with reduced transcription and *de novo* methylation in Rec L cells (Fig. 3A, B).

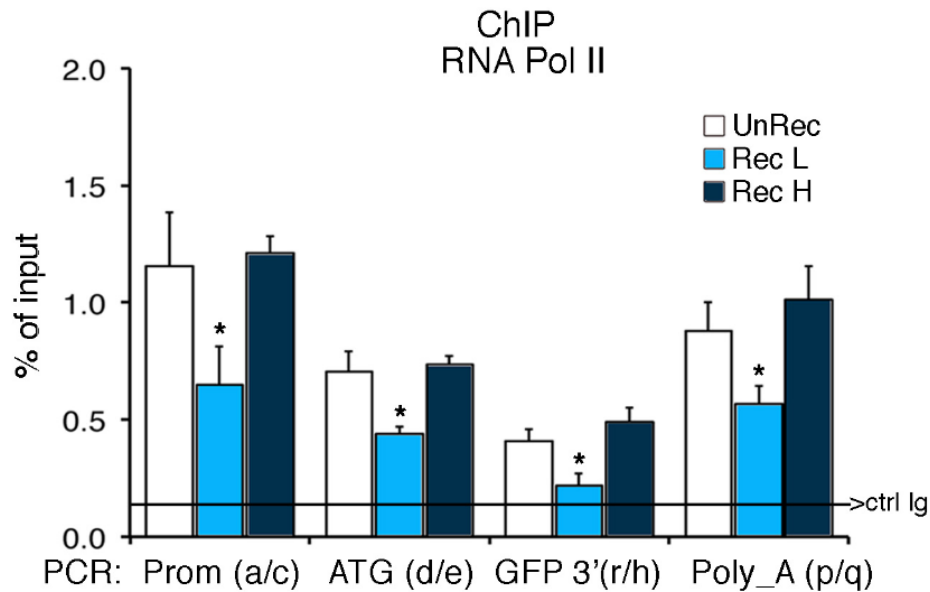
To further characterize loops *A* and *C*, formaldehyde-fixed chromatin was extracted from Rec H and Rec L cells or from cells expressing I-SceI for 48 h and treated with RNase H, which cleaves the RNA in RNA:DNA hybrids. Fig. 4B shows that the loop *A* was sensitive, whereas loop *C* was resistant, to RNase H cleavage. The band in Rec H cells corresponding to loop *C* may represent contamination by L cells, since it was resistant to RNase H. These data indicate that the loops *A* and *C* form very early after the DSB, before Rec H and Rec L cells stabilize (Fig. 4C, left panels). Also these data suggest that loop *A* in Rec H cells is held together by an RNA:DNA hybrid that links the 3' polyA site and the 5' end of the gene (see Fig. 4 B,C).



**Figure 4. Chromatin-DNA domains induced by damage-repair.** **A.** Loss of methylation modifies the chromatin loops induced by HR. Non-recombinant or Rec H or Rec L cells were exposed to 5 or 50  $\mu$ M 5-azadC for 24h + 24h in normal medium (Materials & Methods). 3C analysis was carried out as described in Materials & Methods and formation of loops A and C was quantified by qPCR relative to control ligation. The results derive from at least 3 experiments in triplicate. \* $p < 0.01$  as compared to untreated samples. **B.** Loop A and C in formaldehyde-fixed chromatin treated with RNaseH. Sorted recombinant cells were fixed in formaldehyde (1%) and chromatin was prepared as described in Materials & Methods for 3C analysis, except that an aliquot was digested 30 min at 30°C with 100  $\mu$ g/ml RNaseH (DNase and protease-free). At the end of the incubation, 3C was performed as indicated in Materials & Methods. The ligation products were analysed with the primers by qPCR indicated below the histograms (right panels). **C.** On the left, the upper panels show the loops A and C in mass culture of cells expressing (I-SceI) for 48 h or not (UnRec). The fixed chromatin derived from these cells was treated with RNase H as described in panel B. On the right, the panels show the same experiment carried out on chromatin of sorted Rec H and Rec L cells analysed by semiquantitative PCR.

To analyze further the mechanism of transcription attenuation by DNA methylation following HR in Rec L cells, we determined by ChIP analysis the location of RNA polymerase II (Pol II) at various locations in the GFP gene: the promoter, the translation start site (ATG), the region of the DSB and HR (GFP3') and the 3' polyA site. In non-recombinant or GFP+ cells, Pol II concentrations were highest at the GFP promoter and polyA sites (Fig. 5). The levels of RNA Pol II were reduced at all sites in L cells, suggesting that the

new chromatin loop (loop *C* in Fig. 3A and 4B), induced by local methylation, interferes with transcription even at sites not physically contiguous with the *I-SceI* region (see Fig. 3B).



**Figure 5. Recruitment of RNA Polymerase II at the site of GFP gene before and after the DSB and HR.** Occupancy of RNA polymerase II (Pol II) on the GFP in non recombinant and Rec H and Rec L recombinant cells. The chromatin was immunoprecipitated with antibodies directed against RNA Pol II. The panel shows the recruitment of Pol II at the promoter, translation start site [ATG], the region of DSB/HDR [GFP3'] and the 3' polyA site. The primers used are indicated and their location is shown in Fig. 5A. The black, horizontal, line indicates the percent of input from a control CHIP (Ab: non immune serum). The statistical analysis derives from at least three different experiments in triplicate ( $n \geq 9$ ; Mean  $\pm$  SD); \* $p < 0.01$  (matched pairs t test) compared to non recombinant sample.

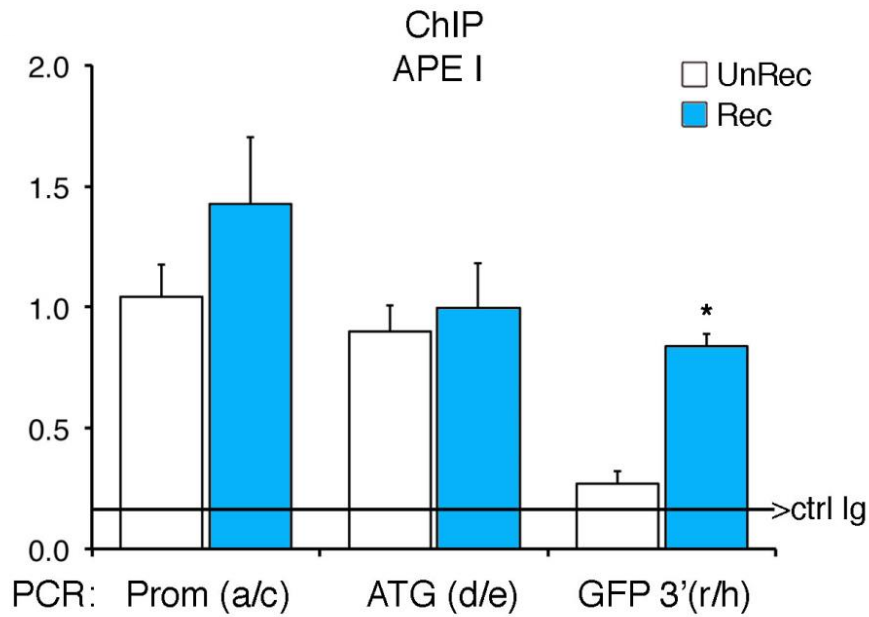
Our data suggest that the initial and transient enrichment of H3K9m3 in the chromatin of damaged DNA is selectively stabilized in Rec L cells by local *de novo* DNA methylation. Methylation of the repaired segment favors the formation and stabilization of loop *C*, which is associated with the reduced GFP transcription of Rec L cells (Fig. 3B and 5). Loop *C*, which is detectable as early as 48 h after the exposure to I-SceI (see Fig. 4C), marks methylated GFP, brings in close proximity the promoter and the polyA site, and facilitates

propagation of the repressive mark, methylated H3K9, from the DSB to the promoter and to the 3' end of the repaired gene(Fig. 4B). Importantly, loss of methylation induced by 5-azadC also reduced 3K9m2/3 markers at these GFP sites. These loci contain few CpGs and are not contiguous on the linear DNA sequence, but are brought in close proximity by loop *C* (Fig. 3B). The appearance of H3K9m3 at these sites is also correlated with the reduction in Rec L cells of RNA Pol II molecules across the repaired GFP gene (Fig.5)

### **4.3. BER enzymes remodel DNA methylation during HR**

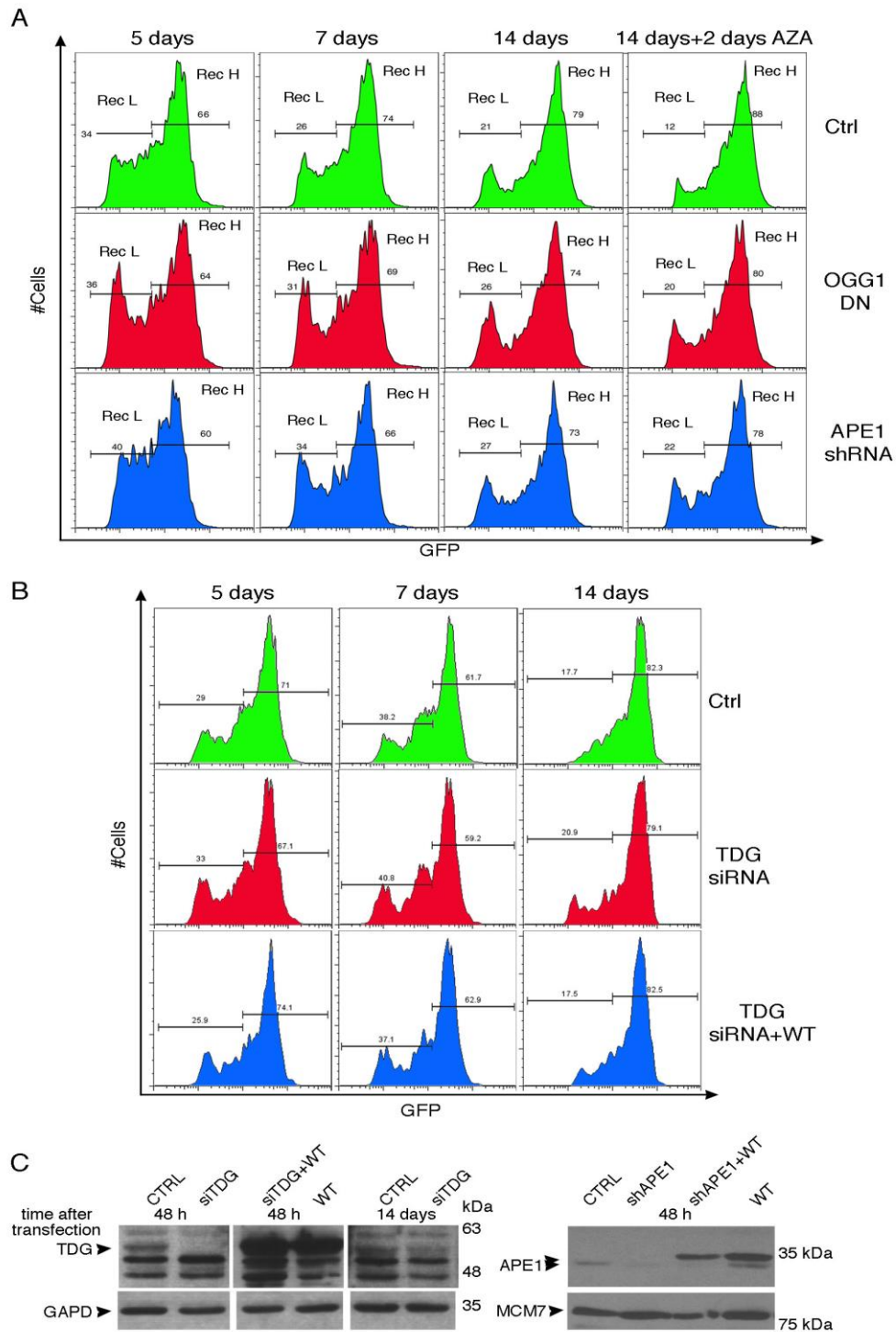
We showed previously that inhibiting transcription after HR in Rec L cells with short pulses of Actinomycin D increased the methylation content of the repaired GFP gene (Morano A. et al 2014). Transcription is associated with active DNA demethylation(Williams, K. et al. 2011). BER and NER enzymes are involved in transcription and repair-induced active demethylation (Zuchegna C.et al. 2014; Perillo B. et al. 2008;Amente S. et al. 2010;Le May N.et al. 2012; Kohli R.M. et al. 2013). We hypothesized that depletion of these enzymes might, like Actinomycin D, inhibit transcription and increase HR-induced methylation. To test this idea, we first monitored recruitment of BER enzymes to repaired chromatin and second, determined the impact of silencing these enzymes during and after repair on DNA methylation and GFP expression. We found that APE1, the BER apurinic site nuclease, was enriched at recombinant chromatin, particularly at the *GFP* 3'end (Fig. 6).





**Figure 6. Recruitment of RNA polymerase at the sites of GFP gene before and after the DSB and HR.** Occupancy of Apurinic/apyrimidinic endonuclease (APE1) on the GFP in non recombinant and recombinant cells. The chromatin was immunoprecipitated with antibodies directed against APE1. The panel shows the recruitment of APE1 at the promoter, translation start site [ATG], the region of DSB/HDR [GFP 3']. The primers used are indicated and their location is shown in Fig. 5A. The black, horizontal, line indicates the percent of input from a control CHIP (Ab: non immune serum). The statistical analysis derives from at least three different experiments in triplicate ( $n \geq 9$ ; Mean  $\pm$  SD); \* $p < 0.01$  (matched pairs t test) compared to non recombinant sample.

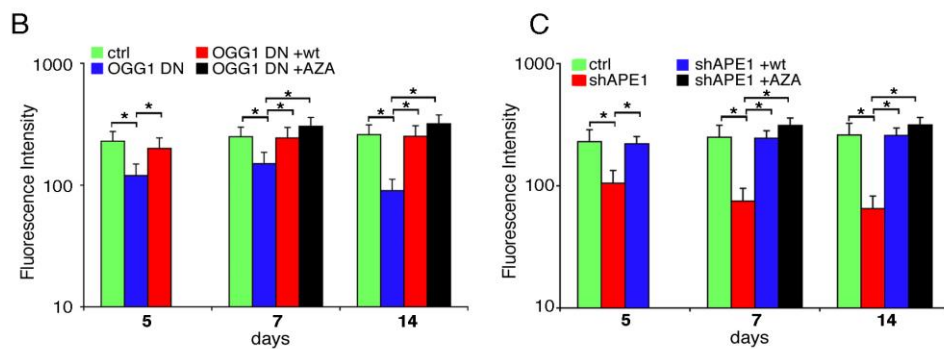
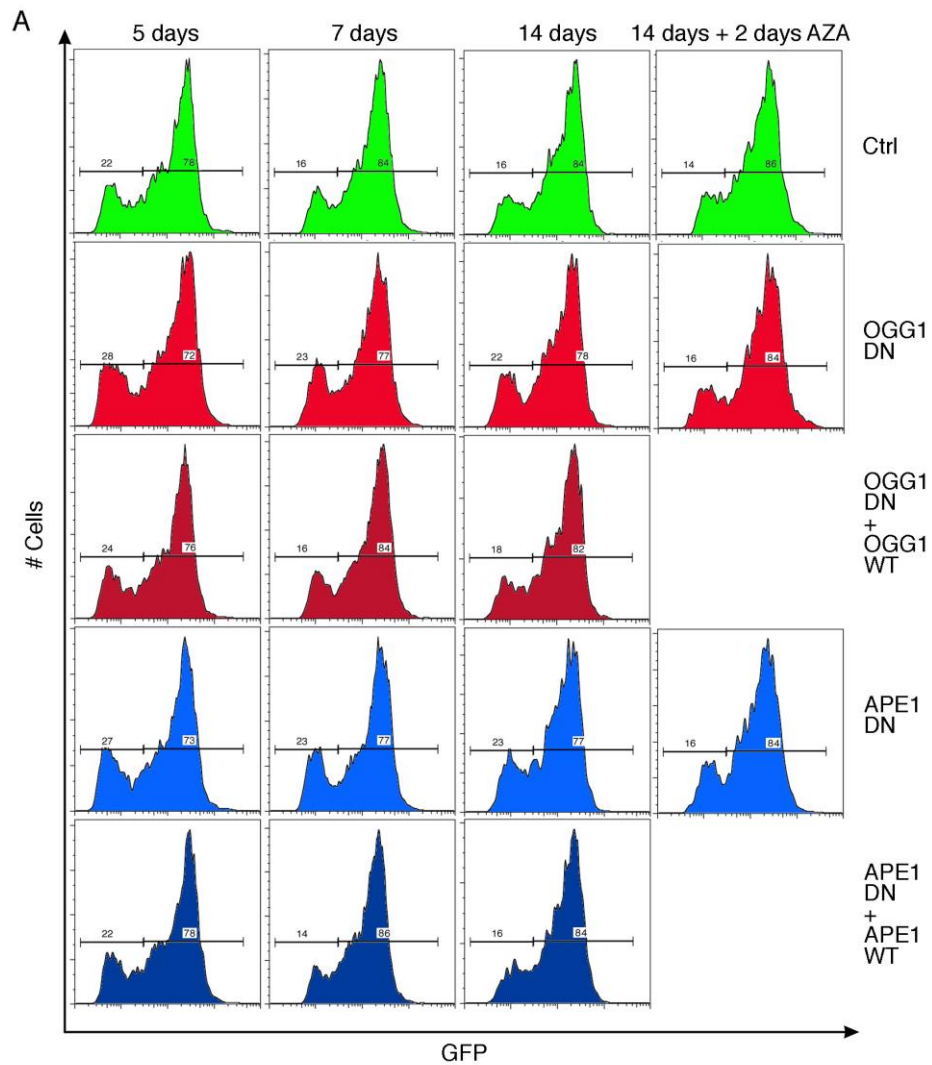
We also inhibited APE1 during repair, as well as two other enzymes involved in BER and transcription, the 8-oxoguanine glycosylase, OGG1, and TDG, which has been directly implicated in active DNA demethylation (Wu S.C. and Zhang Y. 2010; Cortellino S. et al. 2011). The effects on GFP expression were evaluated beginning 5 days after DSB formation, when their mRNA and protein levels had returned to normal after knock-down (Fig. 7C). Inhibition of OGG1 or APE1 with dominant-negative variants (Figs. 7A, 8A and 8B) or APE1 with shRNA (Figs. 7A, 8C) significantly increased the L/H ratio and reduced GFP expression in both L and H cells (Fig. 9).



**Figure 7. Inhibition of BER during repair inhibits transcription and increases methylation of the repaired DNA.** **A.** Cytofluorimetric analysis of GFP positive DRGFP HeLa cells transfected with I-SceI and a dominant negative expression vector for OGG1 or a vector

expressing APE1 shRNA. Endogenous APE1 protein was reduced by 70% at 36h after transfection, GFP expression was measured by FACS at 5,7 and 14 days after transfection as described in Experimental Procedures. 14 days after transfection, cells were treated with 10uM 5-azadC for 24h and analyzed by cytofluorimetry 48h later. Percentage of Rec H and Rec L cells are indicated in each plot. **B.** Cytofluorimetric analysis of GFP positive after HR in TG silenced cells. **C.** TDG and APE1 in silenced cells. Immunoblot analysis of TDG-silenced expression of GFP is reduced (panel B). The right panel shows the immunoblot for APE1 in cells transfected with the shAPE1 and the wild type expression vectors. The 2 bands visible in the blot correspond to the endogenous (lower) and the exogenous (upper) bands.

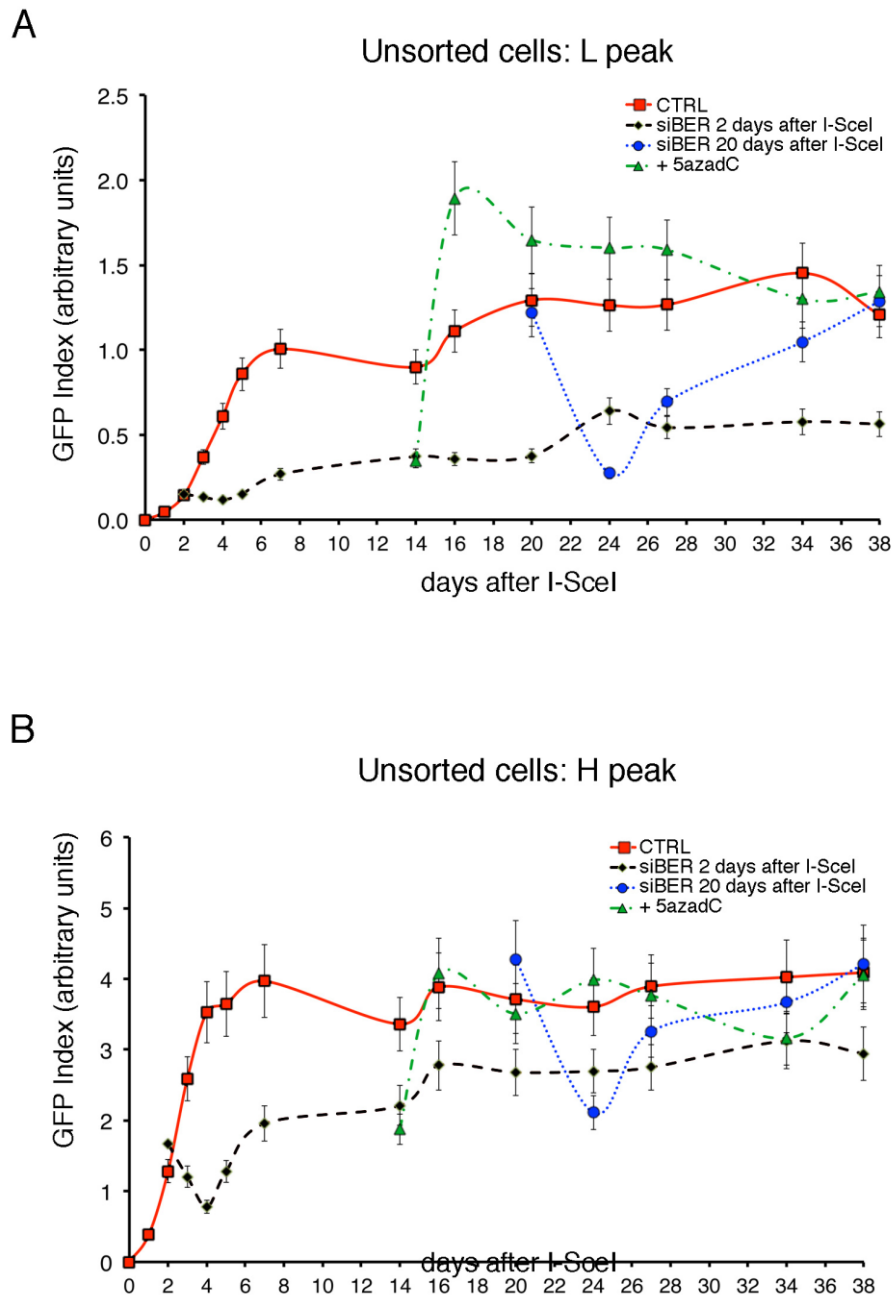
Inhibition of OGG1 or APE1 with dominant-negative variants (Figs. 7A, 8A and 8B) or APE1 with shRNA (Figs. 7A, 8C) significantly increased the L/H ratio and reduced GFP expression in both L and H cells (Fig. 9).



**Figure 8. Rescue of GFP expression in cells in which BER enzyme are inhibited during repair.** *A.* Cytofluorimetric analysis of GFP positive DRGFP HeLa cells co-transfected with I-SceI and dominant negative expression vectors for OGG1 or APE1 or the WT variant (see

Experimental Procedures). GFP expression was measured by FACS at 5, 7 and 14 days after transfection as described in Experimental Procedures. 14 days after transfection, cells were treated with 10 $\mu$ M 5-azadC for 24h and analyzed by cytofluorimetry 48h later. Percentage of Rec H and Rec L cells are indicated in each plot. **B.C.** Statistical analysis of GFP expression (mean of fluorescence intensity) in OGG1 and APE1 depleted cells in independent experiment. The effects on GFP expression of OGG1 and APE1 is shown in panel A. The statistical analysis derives from at least 3 experiment in triplicate ( $n \geq 9$ ; Mean  $\pm$  SD); \* $p < 0.01$  (matched pairs t test).

This reduction was reversed by 5-azadC or by expressing the wild-type versions of the enzymes (Figs. 7A, 8). Cells depleted of thymine-DNA glycosylase (TDG) with siRNA displayed a similar phenotype, which was reversed by expressing the wild-type vectors (Fig. 7B). The reduction of GFP expression and the reversion by 5-azadC were stable and permanent up to 38 days after exposure to I-SceI (Fig. 9). We also silenced BER enzymes 20 days after DSB formation, when repair was complete (Fig. 9). At this time, both Rec L and Rec H cells displayed reduced GFP levels when APE1 and TDG were silenced, but the inhibition of GFP expression disappeared when the silenced protein levels returned to pre-treatment levels. Figure 9 shows a summary of the results derived from many (~11) independent experiments performed within a 38-day time frame after exposure to I-SceI.



**Figure 9. Inhibition of BER during repair reduces transcription and increases methylation of the repaired DNA. A and B.** Summary of the cytofluorimetric analysis of the total cell population exposed to I-SceI from day 0 to 38. On the ordinate is shown an arbitrary unit, the GFP index, i.e. the product of GFP intensity and the reciprocal fraction of cells in the fluorescence gate. This value normalizes the frequency of recombination with the intensity of the signal. The upper and low panels show the GFP index in the higher (H peak) and lower (L

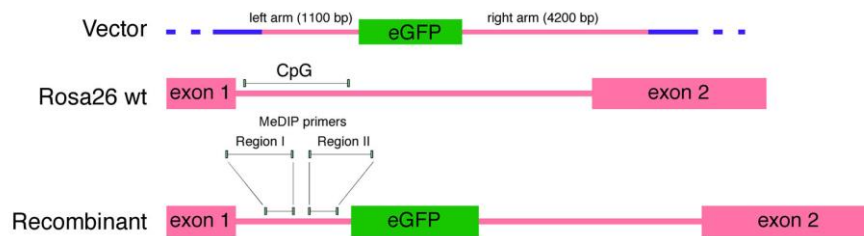
peak) expressors, respectively. The red boxes represent the mean values of the GFP index found in cells transfected with scramble siRNA at day 2, or 20 or untransfected cells. Each value is compared to its matched controls, i.e. cells untransfected or transfected with scrambled siRNA and analyzed at the same time points of cells transfected with siBER. siBER did not modify the number of GFP recombinants. The cells were divided into 2 aliquots: at day 2 cells were exposed to si-shBER RNAs (shAPE1-siTDG1) and cytofluorimetry was performed on the days indicated in the graph. The other aliquot of cells was treated with the same reagents 20 days after transfection and the cytofluorimetric analysis was performed as indicated in the graph (blue broken lines). At day 14, 5-azadC (10 µg/ml) was added for 2 days, and cytofluorimetry was recorded as indicated by the green broken line. The specific data are shown in Fig. 7 and 8.

Collectively, these data show that silencing of BER-NER enzymes permanently modified GFP expression in Rec L and Rec H cells only when it occurs during repair, whereas the changes of GFP expression were transient if silencing is performed 20 days after the repair (Fig. 9). BER depletion induces transient inhibition of expression of many genes in addition to GFP (Zucchegna C. et al.2014; and data not shown).

Similarly, inhibition of transcription by Actinomycin D permanently reduces GFP expression only if added shortly after the repair period; the effect of the drug is transient if administered 21 days after repair (Morano A. et al. 2014). Together, these results indicate that transcription inhibition increases DNA methylation only during repair of DNA damage. Importantly, Fig. 9 shows that there is a precise time window set by DNA damage and repair, in which inhibition of transcription or of repair associated with transcription, lead to stable methylation. We suggest that this period is required for stabilization of novel chromatin domains induced by local DNA methylation after HR (Figs. 1 and 2).

#### 4.4. Homologous targeting of GFP in ES cells generates clones with various levels of GFP expression and DNA methylation

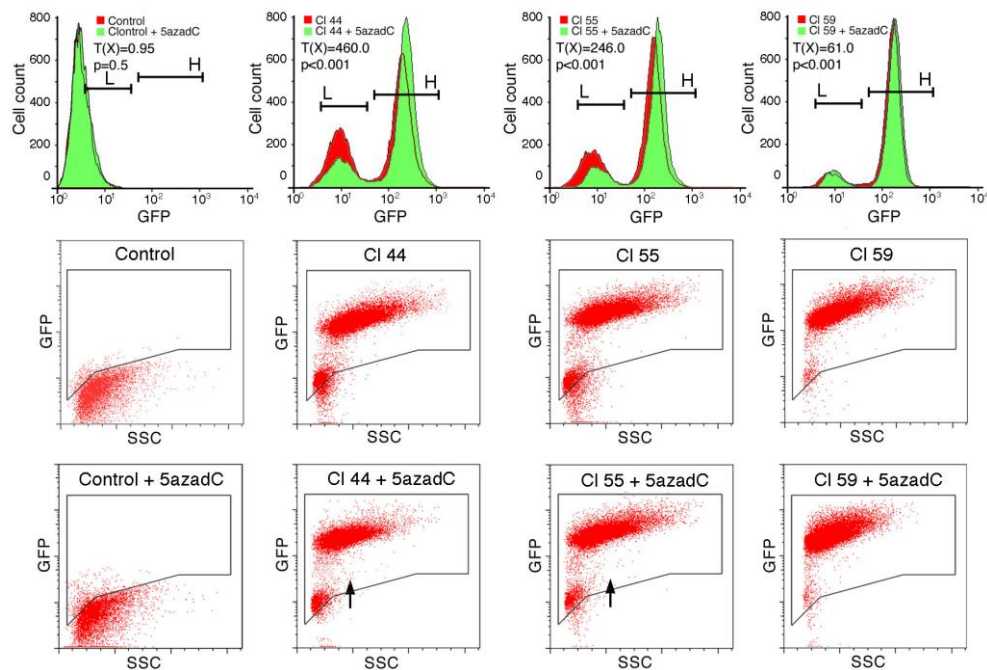
The data presented above indicate that both HR and transcription contribute to the final and stable methylation status of the repaired DNA segment. Our experimental system relies on an artificially generated DSB by the meganuclease, I-SceI. To create a DSB by other means, we targeted GFP to a DNA locus by homologous recombination (Fig. 10). Integration requires formation of a DSB and should thus generate both Rec H and Rec L cells. Specifically, we predicted that the inserted gene will have different expression levels in different, although genetically identical, clones. To this end we isolated 3 ES clones carrying a single copy of CMV<sub>1</sub>EGFP targeted to the Rosa26 locus in the mouse genome within seven days after transfection (A. Simeone and D. Acampora, unpublished observations and Fig. 10).



**Figure 10. Rosa 26 locus and targeted locus.** The panel shows: 1. the structure of the targeting vector with the two homologous regions; 2. the structure of the mouse Rosa26 locus and; 3. the structure of the targeted locus. A segment at the 5' end, containing a CpGs island, is shown (arrows). The primers for MEDIP analysis are located in regions I and II.

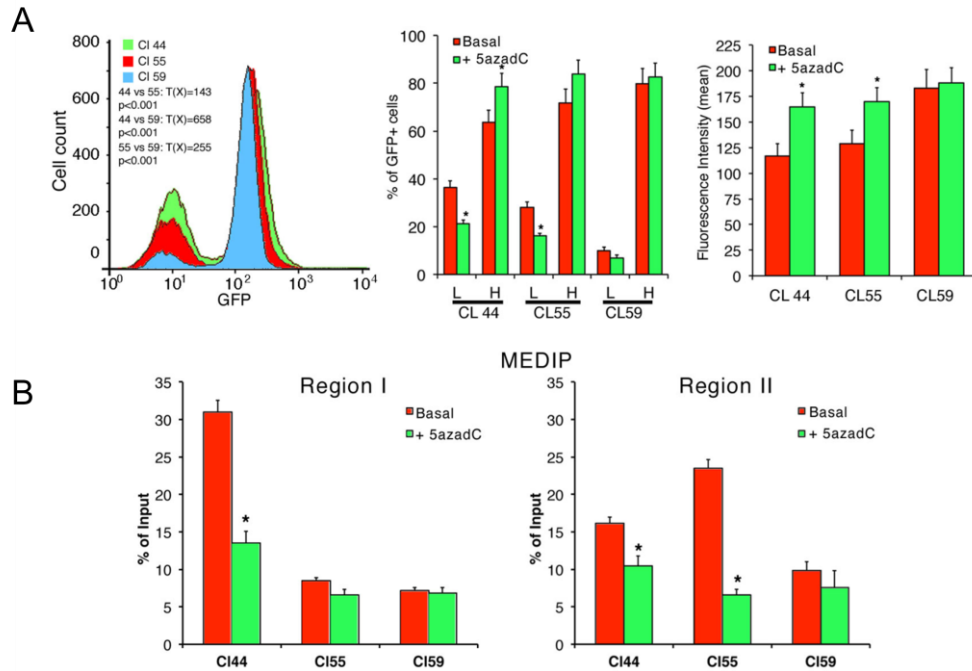
These clones, 44, 55 and 59 were characterized for GFP expression. Clones 44 and 55 contain two populations that vary in GFP expression levels, whereas clone 59 contains essentially only Rec H-like cells (Fig. 11).





**Figure 11.** Cytofluorimetric analysis of the 3 clones exposed or not to 0.5  $\mu$ M 5-azadC for 4 days and analyzed 48 h later. Dot Plot scans are shown to illustrate the composition of GFP + or GFP – cells. The arrows indicate the shift of the L population after 5-azadC treatment. Differences in GFP expression between control and 5-azadC treated cells were tested for statistical significance using the Chi Square test, T(X), (Population Comparison module of the FlowJo software from Tree Star). Cl 44, untreated vs 5-azadC T(X)=460, p>0.001; Cl 55, untreated vs 5-azadC T(X)=246, p>0.001; Cl 59, untreated vs 5-azadC T(X)=61, p<0.001.

We propose that clones 44 and 55 are similar to Rec H and Rec L clones found in DR-GFP HeLa cells and clone 59 is similar to Rec H clones containing essentially hypomethylated GFP. To test this hypothesis, we treated the cells with 5-azadC and measured GFP expression. Figure 12A shows that treatment with 5-azadC shifted Rec H cells to the right (higher expression) and reduced the number of Rec L cells in clones 44 and 55, but not in clone 59 (Fig. 11 and 12A).



**Figure 12. Homologous targeting of GFP to the Rosa26 locus generates clones with variable expression and DNA methylation.** Panel A shows the overlapping profiles of the three clones without treatment to compare the relative GFP expression levels. CI 44 vs CI55 T(X)=143, p<0.001; CI 44 vs CI59 T(X)=658, p>0.001; CI 55 vs CI59 T(X)=255, p>0.001. On the right of the panel quantitative analysis of GFP expression in H and L cells before or after 5-azadC treatment (% of GFP+ cells and mean of fluorescence intensity, respectively) are shown. Differences between treatments were tested for statistical significance using Student's matched pairs *t* test: \*p<0.001, Panel B shows MEDIP analysis of region I and II, respectively in the 3 clones. \*p<0.001.

We also carried out a DNA methylation analysis of the CpGisland present at the 5' end of the homologous targeting sequence, with primers specific to subregions I and II (Fig. 10). MEDIP analysis shows that region I in clone 44, and region II in clone 55 were differentially methylated compared to clone 59 (Fig. 12B). These methylation differences were reversed by 5-azadC (Fig. 12B). The localization and the clone specificity of methylation suggest that a DSB occurred upstream of region I in clone 44 and between the regions I and II in clone 55 during homologous pairing. The 3' end of this hypothetical DSB (along the direction of transcription) was methylated following gene targeting.

These data extend the notion of HR induced-methylation and suggest a general mechanism that modifies expression of targeted genes by homologous recombination.

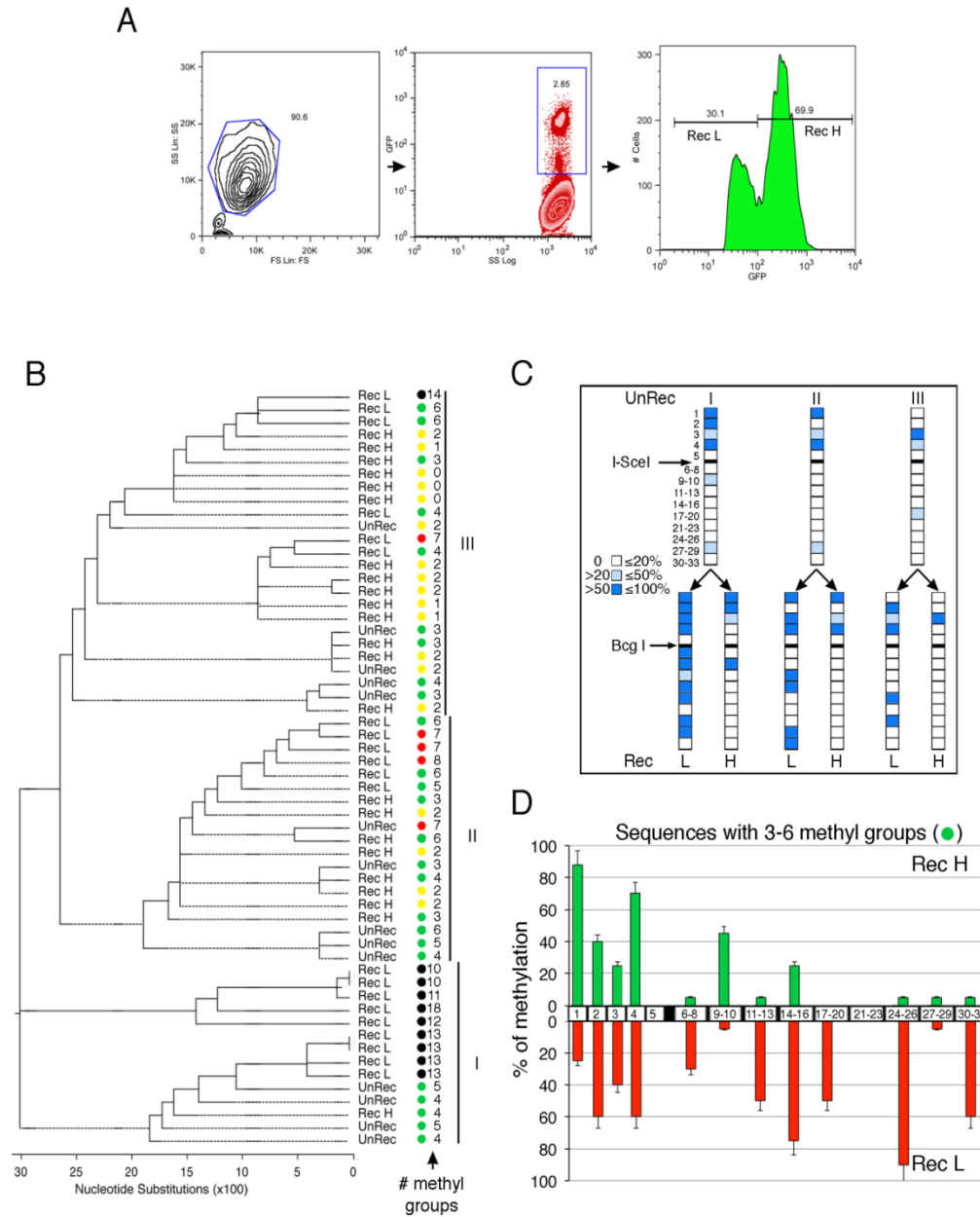
#### **4.5. Discrete DNA methylation patterns mark clones with distinct gene expression levels**

Our data indicate that DNA methylation status is highly dynamic, and can be reshaped to various extents during and after DNA damage-repair. The effect of DNA methylation *per se* on gene expression can be highly variable, because the progeny of a single cell displays different DNA methylation profiles that are related to GFP expression levels. Over time, the DNA methylation status of Rec H and Rec L cells stabilizes and generates cells with different but heritable GFP expression levels, characterized by specific GFP chromatin domain patterns (see Figs. 1 and 2).

To find the relation between methylation and gene expression, we compared the location and the number of methylated CpGs with GFP expression levels in the most frequent GFP molecules isolated from sorted Rec H and Rec L cells. We ordered the methylated GFP in families sharing mCpGs at identical locations to define *epigenetic haplotypes* and we asked if molecules with the same number of mCpGs, but localized in different sites in the gene (i.e., different haplotypes), express similar levels of GFP.

Clustering analysis identified three main groups of molecules, each containing UnRec, Rec H and RecL molecules, sharing a similar methylation pattern at the 5' end of the DSB (Fig. 13B). This 5' end region contains 4 CpGs, whose methylation status is stable and is not affected by HR, i.e., is the same in UnRec (cells not exposed to I-SceI) and in Rec cells. Methylation of the 3' end region, instead is significantly modified by HR (Morano A. et al. 2014). Figure 13 shows that each group gave rise to Rec L and Rec H cells

whose methylation pattern is related to UnRec cells, but has been modified by HR-induced methylation. The most predictive methylation pattern associated with expression is located in the region at the 3' end of the DSB, where hypermethylation marks Rec L and hypomethylation marks Rec H clones. In all groups, there are molecules with an intermediate level of methylation (3-6 methyl groups/mol. indicated as green circles in 13B), in which the position of methylated Cs is relevant for the expression. Figure 13D shows the relative methylation frequency of CpGs in this segment modified by HR. For example, CpGs 11-13 and 17-20 are methylated in Rec L cells and never in Rec H clones, even in those clones carrying the same number of methylCpGs found in Rec L (Fig. 13D, red and green columns).



**Figure13. Tracking methylated alleles of the GFP locus after HR.** **A.** Cytofluorimetric analysis of Rec H and Rec L cells, as described in Materials & Methods. **B.** Epiallele analysis of 58 representative GFP DNA molecules before and after HR. These molecules were derived from a collection of 1000 sequences of GFP bisulfite DNA before (UnRec) and after (Rec) HR and are represented at least 10 times in the cell population (A. Morano et al., 2014). The gene ontology dendrogram was generated on the basis of C-T (mCpG) localization by clustering analysis (ClustalW). DNA primary sequence was the same for all 58 molecules including non-recombinant GFP, in which the I-SceI sequence was *in silico* converted to BcgI to compare the molecules only on the basis of the methylation profile. Each molecule was characterized according to: 1. the 5' end corresponding to the first 4 CpGs 5' to the I-SceI site (UnRec or Rec H or Rec L) and 2. the 3' end corresponding to the 20 CpGs downstream to the site; 3. the number of methylated CpGs. Molecules with the same or similar number of methylated

CpGs are indicated by circles of the same color (yellow, 0-2 ; green, 3-6; red, 7-8; black, 9 or more). **C.** Schematic diagram showing the groups I, II or III, which include GFP molecules containing CpGs methylated at the same position at 5' and 3' end of the DSB (arrow indicates the DSB) in UnRecombinant, UnRec, (the upper part showing 3 chromosomes) and Recombinant, Rec, (lower part showing 6 chromosomes). Since HR converts the I-SceI site in Bcg I, recombinant molecules are shown with Bcg I site. The color code indicates the frequency of each methylated **C.D.** Intermediate methylated GFP molecules, indicated by the green circles in B, were analysed and the frequency of methylation of CpGs at the 5' and 3' end of DSB (black box) is shown as histogram in Rec H (green) and Rec L (red) clones.

Collectively, these data show that expression of GFP gene is influenced by the methylation status, which is induced by damage repair and modified by ongoing transcription. The total content of methylated CpGs seems not as relevant as the location of the methylated base. In intermediate methylated GFP molecules, the qualitative (location) rather than the total content of methylated Cs is associated with the variation of GFP expression. Since GFP gene expression does not alter the fitness of the cells, the frequency of methylated GFP alleles is constant.

## 5. DISCUSSION

### 5.1. Transient and stable chromatin changes induced by DNA damage and repair.

We have explored the changes of chromatin that accompany DNA methylation at the site of DNA repair. We have demonstrated the histone code change at the DSB site 24h and 48 h after I-SceI expression.

Early after DSB formation, chromatin near the lesion becomes enriched with H3K9m2/3. This modification is essential to recruit the histone acetyl transferase TIP60, which acetylates and stimulates ATM at the site of damage (Sun Y. et al. 2010). Concurrently, DSB-carrying cells transiently lose H3K4 methylation, and regain it 2 to 7 days after DSB. Since the H3K9m2/3 changes occur early after the DSB, when the repair is not yet detectable, we suggest that these changes are induced by DSB and not by the repair. However, further purification of HR cells reveals that H3K9m2/3 is selectively retained after HR only by Rec L clones (Fig. 1A and 1B). Our experiments suggest that maintenance not the establishment of this repressive chromatin mark is secondary to DNA methylation. Thus, treatment of Rec L cells with the DNA demethylating drug, 5-azadC, significantly reduced the levels of H3K9m2/3 on GFP chromatin (Fig. 2). *De novo* methylation of the repaired segment was also responsible for the stabilization of the chromatin loop specific to Rec L cells (loop C in Fig. 3). In Rec L cells, the repressive methylation H3K9m2/3 marks were also present at locations physically distant from the repaired DSB. These sites were, in fact, juxtaposed and linked by loop C, which connects the transcription start and the polyA sites (Figs. 1 and 2). At other physically distant sites (e.g., the puromycin-resistance gene), histone marks were not modified by damage and repair (Cuozzo C. et al. 2007; Morano A. et al 2014; and data not shown).

We suggest that the initial functionally relevant event is the accumulation of H3K9me3 derived from methylation of H3K9m2 induced at the DSB by the recruited histone methyltransferase SUV39, which does not directly influence repair (Morano A. et al. 2014). The increase of this repressive mark contributes to the repression of local transcription induced by DNA damage. Importantly, the local chromatin changes are quite dramatic: loop C, which marks a low-expression fraction of stabilized HR cells, is formed very early and is detectable in the mass culture as soon as 48 h after DSB formation, before repair is complete and the Rec L clones are stabilized (Fig. 4C). Loop A, which connects the polyA site with the CMV promoter, is characteristic of high-expressors. In Rec H or non-recombinant cells, the H3K9m2/3 mark is rapidly lost, probably because is not stabilized by HR induced DNA methylation (Fig. 2). The changes we have described indicate that the DNA and chromatin undergoing HR are marked epigenetically very early during the repair process, and that some of these marks are permanently maintained. Surprisingly, we find evidence that RNA plays a role in determining the pattern of chromatin looping. Thus, loop A is sensitive to RNase H, whereas loop C is increased by enzyme treatment; possibly the A and C loops are in competition. This suggests that RNA is a structural component of the A loop and may differentiate a chromatin loop associated with active transcription (Rec H clones) from a loop associated with reduced transcription (Rec L clones; Fig. 4B). Although the nature of the RNA is unknown, sensitivity to RNase H suggests that the integrity of the loop is maintained by a DNA:RNA hybrid, presumably at the base of the loop. The hybridizing RNA might arise from a rare splicing event, e.g. trans-splicing (for review see Fang W. and Landweber L.F. 2013), which fused the 5' and 3' ends of the GFP transcript. Characterization of the putative fused RNA is currently under active investigation.



## 5.2. BER enzymes remodel DNA methylation during HR

Demethylation of DNA has recently been linked to BER enzymes, which remove mismatched or alkylated bases. 5mC can be removed directly by specific glycosylases. Direct removal of 5mC has been shown in plants, which express 5 different types of 5mC specific glycosylases (Zhu J.K. 2009). Thus far, no 5mC specific glycosylase has been found in mammals. In mammals, demethylation requires that 5mC is first modified by deaminases (AID or APOBEC, Morgan et al. 2004) or by hydroxylation by TET enzymes (Tahiliani et al. 2009), and then, removed by Thymine DNAglycosylase (TDG).

In literature has been demonstrated that GADD45A binds DSB and inhibits *de novo* methylation induced by HDR. The effect on GFP methylation is likewise transient and appears to be due to inhibition of methylation, rather than active demethylation linked to transcription (Morano A. et al. 2014). After repair, DNA methylation is not static but is progressively re- modeled. The actors of this second phase, APE1 and OGG1, are involved in BER and are also required for transcription (Perillo et al., 2008; Bhakat et al. 2009; Amente et al., 2010). APE1 is the apurinic site endonuclease, which repairs apurinic/apyrimidinic sites (AP) generated after excision of oxidized bases (Bhakat et al., 2009). OGG1, is the 8- oxoG specific glycosylase, which is involved in repair of oxidized G and probably pairs with oxidized 5mC during transcription. Our data indicate that BER (TDG, APE1 and OGG1) enzymes are recruited to the DSB and remain on the site after repair (Figs. 6,7,8). Silencing or inactivation of APE1 or OGG1 enhances methylation and reduces expression of GFP (Fig. 7). We propose that BER enzymes, loaded by the transcription machinery(Wu & Zhang, 2010) contribute to de methylation of L cells and their conversion to H cells.

### **5.3. Relation between methylation profiles and levels of gene expression**

We demonstrated that in the absence of selection (such as is the case for GFP expression), we can isolate recombinant clones carrying distinct methylation profiles of the repaired gene. Comparing the levels of expression of GFP with the methylation profiles of single DNA molecules, we find a relation between the profile and the expression of the repaired gene. We describe clones carrying the same number of methyl groups (Fig. 3), but displaying different levels of GFP expression. These differences are erased by 5-azadC treatment, indicating that the variations are DNA methylation-dependent. We find that the differences instead reflect the methylation state of specific CpGs (Figs. 1, 2 and ref. Morano A. et al. 2014). We propose that certain methylated CpGs may reduce gene expression by stabilizing repressive chromatin loops (see Figs. 2 and 13). The variability of GFP expression is affected by editing of local methylation by transcription or by active demethylation (Fig. 3 and ref. Wu S.C. and Zhang Y. 2010). In principle, the variability in methylation profiles in a complex cell population could enhance adaptation of the population to environmental differences. Since the methylation profiles of Rec H and Rec L clones are stable and inheritable, we exploited the variability of these profiles to trace specific epialleles generated during HR. We applied gene ontology analysis to methylation, and generated haplotypes of linked methylated CpGs. In conclusion, our data demonstrate that DNA methylation induced by repair substantially contributes to differential expression of the repaired gene by stably modifying chromatin and DNA. Furthermore, our data opens a way to decode the history of cellular evolution through qualitative analysis of methylation.

## 6. CONCLUSIONS

Using a well-defined genetic system, we demonstrated that DNA damage (DSB) and the repair through homologous recombination leave methylation scars (in the form of cytosine methylation, a DNA base) on one strand of the repaired segment.

These are epigenetic “scars”, that do not alter the primary genetic information, but reduce the expression of the gene repaired. This mechanism may be very important to adapt the cells to new environments. For example, the damaged cell after repair generates two daughter cells: one in which the repaired DNA segment is methylated and the other, in which the same DNA segment is not methylated. If the expression of the repaired gene is detrimental for growth or survival, the cell with the methylated gene is positively selected (see video Morano A. et al. 2014).

On the basis of this information gathered on the mechanism of homologous directed repair –induced methylation, somatic methylated epialleles are stably inherited and in the absence of selection are randomly distributed in cell populations. If HDR-induced methylation occurs in a suppressor gene, the cell carrying the methylated epialleles grows faster than other cells in the population. Proliferation of clones in a complex mixture of cells should, at least in principle, select specific epialleles, which expression or silencing favour growth, independently of the primary cause of clonal expansion. Combining qualitative (location of methylated CpGs in the same molecule) with quantitative (how many times the specific CpG is found methylated), we were able to define and monitor the evolution of specific suppressor gene epialleles during neoplastic progression and in some cases to anticipate the chemoresistance.

## **7. ACKNOWLEDGEMENTS**

At the conclusion of my PhD program, I would like to thank...

My Supervisor and PhD Coordinator, professor Vittorio Enrico Avvedimento, for his contribution to the growth of my preparation and for conveying to me the profound passion for research.

The Company Microtech that, with the fellowship "Doctorate in Industry" POR Campania FSE 2007-2013, partly supported the research during my PhD.

Professor Antonio Porcellini, for the countless tips and for the support during my PhD.

Dr. Antonio Pezone, life partner and colleague, for the love, support, patience and encouragement, thanks to which, I was able to overcome any obstacle encountered during my PhD, who is sharing with him the most beautiful gift that life can give us.

Dr. Rosaria Landi, for the active collaboration during my PhD.

All students and colleagues from my laboratory for showing kindness.

My parents, Raffaele and Anna, and my brother, Fernando, for their love and for conveying to me the perseverance to achieve the targets.

My husband's parents, Vincenzo and Giuseppina, and my brothers in law, for their love and for supporting me in every moment of my training path.

## 8. REFERENCES

- Amente S., Bertoni A., Morano A., Lania L., Avvedimento E.V., and Majello B. *LSD1-mediated demethylation of histone H3 lysine 4 triggers Myc induced transcription.* (2010). *Oncogene* **29**, 3691-3702
- Bhakat K.K., Mokkapati S.K., Boldogh I., Hazra T.K., Mitra S. *Acetylation of human 8-oxoguanine-DNA glycosylase by p300 and its role in 8-oxoguanine repair in vivo.* (2006). *Mol Cell Biol***26**, 1654-1656.
- Bhakat K.K., Mantha A.K., Mitra S. *Transcriptional Regulatory Functions of Mammalian AP-endonuclease (APE1/Ref-1), an Essential Multifunctional Protein.* (2009).*Antioxid Redox Signal*.**11**, 621–638.
- Baylin S.B., Herman J.G., Graff J.R., Vertino P.M. and Issa J.P. *Alterations in DNA methylation: a fundamental aspect of neoplasia.* (1998). *Adv Cancer Res* **72**, 141–196.
- Bird A.P. e Wolffe A.P. *Methylation-induced repression-belts, braces, and chromatin.* (1999). *Cell* **99** (5), 451-454.
- Burma S., Chen B.P., Chen D.J. *Role of non-homologous end joining (NHEJ) in maintaining genomic integrity.*(2006). *DNA Repair***5**,1042-1048.
- Cortellino S., XuJ., Sannai M., Moore R., Caretti E., Cigliano A., Le Coz M., Devarajan K., Wessels A., Soprano D. et al. *Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination base excision repair.* (2011). *Cell* **146**, 67-79.
- Cromie G.A., Connelley J.C. and Leach D.R.F. *Recombination at double-strand breaks and DNA ends: conserved mechanism from phage to human.*(2001). *Moll. Cell*.**8**, 1163-1174.
- Cuozzo C., Porcellini A., AngrisanoT., Morano A., Lee B., Di Pardo A.,Messina S., Iuliano R., Fusco A., Santillo M.R., Muller M.T., Chiariotti L.,Gottesman M.E. and Avvedimento E.V.*DNA Damage, Homology-DirectedRepair,and DNA Methylation.*(2007). *PLoS Genetics***3** (7), 1144-1162.
- Eden S. e Cedar H. *Role of DNA methylation in the regulation of transcription.*(1994). *Curr. Opinion Genet. & Dev.* **4**, 255-259.
- Essers J., Van Steeg H., De Wit J., Swagemarkers S.M.A., Vermeij M.,Hoeijmakers J.H.J. and Kanaar R. *Homologous and non-homologous*

*recombination differentially affect DNA damage repair in mice.* (2000). *EMBO J.***19**, 1703-1710.

Fang W., Landweber L.F. *RNA-mediated genome rearrangement: hypotheses and evidence.* (2013). *Bioessays.***35**, 84-87.

Ha K., Lee G.E., Pali S.S., Brown K.D., Takeda Y., Liu K., Bhalla K.N. and Robertson K.D. *Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery.* (2011). *Hum. Mol. Genet.***20**, 126–140.

Harper J.W., Elledge S.J. *The DNA damage response: ten years after.* (2007). *Mol Cell.***28**, 739–745.

Hoeijmakers J.H.J. *Genome maintenance mechanisms for preventing cancer. A highly informative review of the links between DNA damage, DNA repair pathways and their defects contributing to tumorigenesis.* (2001). *Nature.***411**, 366–374.

Hong L., Suk P., Masaki Okano, Ruth Juttermann, Kendrick A. Goss, Rudolf Jaenisch and En Li. *De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells.* (1996). *Development.***122**, 3195-3205

Jasin M. *Genetic manipulation of genomes with rare-cutting endonucleases.* (1996). *Trends Genet.***12**, 224-228.

Jiricny J. *The multifaceted mismatch-repair system.* (2006). *Nat Rev Mol Cell Biol.* **7**, 335–346.

Kawanishi S., Hiraku Y., Pinlaor S., Ma N. *Oxidative and nitrative DNA damage in animals and patients with inflammatory diseases in relation to inflammation-related carcinogenesis.* (2006). *Biol Chem.* **387**, 365–372.

Kohli R.M., Zhang Y. *TET enzymes, TDG and the dynamics of DNA demethylation.* (2013). *Nature* **502**, 472-479. Review.

Laird P.W., Jackson-Grusby L., Fazeli A., Dickinson S.L., Jung W.E., Weinberg R.A. e Jaenisch R. *Suppression of intestinal neoplasia by DNA hypomethylation.* (1995). *Cell* **81**, 197-205.

Lee B., Morano A., Porcellini A., Muller M.T. *GADD45 $\alpha$  inhibition of DNMT1 dependent DNA methylation during homology directed DNA repair.* (2011). *Nucleic Acids Res.* **40**, 2481-2493

Le May N., Fradin D., Iltis I., Bougneres P. and Egly J.M. *XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes.* (2012). *Mol Cell*.**47**, 622-632.

Li E., Bestor T.H. and Jaenisch R. *Targeted mutation of the DNAmethyltransferase gene results in embryonic lethality.* (1992). *Cell***69**, 915–926.

Li E., Beard C. & Jaenisch R. *Role for DNA methylation in genomic imprinting.* (1993). *Nature***366**, 362–365.

Lindahl T., Barnes D.E. *Repair of endogenous DNA damage. An excellent overview of the extent of endogenous DNA damage, the types of DNA lesions arising from cell autonomous sources, and the pathways that repair such lesions.* (2000). *Cold Spring Harb Symp Quant Biol.* **65**, 127–133.

Loeb L.A., Monnat R.J. Jr. *DNA polymerases and human disease.* (2008). *Nat Rev Genet.***9**, 594–604.

McCabe M.T., Low J.A., Daignault S., Imperiale M.J., Wojno K.J. and Day ML. *Inhibition of DNA methyltransferase activity prevents tumorigenesis in a mouse model of prostate cancer.*(2006). *Cancer Res* **66**, 385–392.

Morano A., Angrisano T., Russo G., Landi R., Pezone A., Bartollino S., Zuchegna C., Babbio F., Bonapace I.M., Allen B., et al.. *Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene.* (2014). *Nucleic Acids Res.* **42**, 804–821.

Morgan H.D., Dean W., Coker H.A., Reik W., Petersen-Mahrt S.K.. *Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming.* (2004) *J Biol Chem.* **10**; 279(50), 52353-60.

O'Hagan H.M., Mohammad H.P. and Baylin S.B. *Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island.* (2008). *PLoS Genet.*,**4**, e1000155.

Panning B. e Jaenisch R. *RNA and epigenetic regulation of X chromosome inactivation.* (1998). *Cell* **93**, 305-308.

Perillo B., Ombra M.N., Bertoni A., Cuzzo C., Sacchetti S., Sasso A., Chiariotti L., Malorni A., Abbondanza C., and Avvedimento E.V. *DNA*

*oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression.* (2008). *Science* **319**, 202-206.16.

Pierce A.J., Johnson R.D., Thompson L.H. and Jasin M. *XRCC3 promotes homology-directed repair of DNA damage in mammalian cells.* (1999). *Genes Dev.* **13**, 2633–2638.

Rountree M.R. and Selker E.U. *DNA methylation inhibits elongation but not initiation of transcription in Neurospora crassa.* (1997) *Genes Dev.***11**, 2383–2395.

Sandovici I., Kassovska-Bratinova S., Vaughan J.E., Stewart R., Leppert M. and Sapienza C. *Human imprinted chromosomal regions are historical hot-spots of recombination.* (2006). *PLoS Genet.***2**, e101.

Santoyo G. and Romero D. *Gene conversion and concerted evolution in bacterial genomes.* (2005). *FEMS Micro.Rev.* **29**, 169–183.

Sigurdsson M.I., Smith A.V., Bjornsson H.T. and Jonsson J.J. *HapMap methylation-associated SNPs markers of germline DNA methylation positively correlate with regional levels of human meiotic recombination.* (2009). *Genome Res.*, **19**, 581–589.

Stewart, M.D., Li, J., Wong, J. *Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment.* (2005). *Mol Cell Biol.* **25**, 2525-2538

Sun Y., Jiang X., Xu Y., Ayrapetov M.K., Moreau L.A., Whetstine J.R., Price B.D. *Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60.* (2009). *Nat Cell Biol.* **11**, 1376-1382.

Sun, Y., Jiang, X., Price, B.D. *Tip60: connecting chromatin to DNA damage signaling.* (2010). *Cell Cycle* **9**, 930-936.

Tahiliani M., Koh K.P., Shen Y., Pastor W.A., Bandukwala H., Brudno Y., Agarwal S., Iyer L.M., Liu D.R., Aravind L., Rao A. *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1.* (2009). *Science.* **15**; 324 (5929) :930-5.

Vaissiere T., Herceg Z. *Histone code in the cross-talk during DNA damage signaling.* (2010). *Cell Res.* **20**, 113-115.



- Valko M., Rhodes C.J., Moncol J., Izakovic M., Mazur M. *Free radicals, metals and antioxidants in oxidative stress-induced cancer.*(2006). *Chem Biol Interact.* **160**, 1–40.
- Vascotto C., Fantini D., Romanello M., Cesaratto L., Deganuto M. et al. *APE1/Ref-1 interacts with NPM1 within nucleoli and plays a role in the rRNA quality control process.* (2009). *Mol Cell Biol***29**, 1834-1854.
- Yan P.S., Shi H., Rahmatpanah F., Hsiau T.H., Hsiau A.H., Leu Y.W. et al. *Differential distribution of DNA methylation within the RASSF1A CpG island in breast cancer.* (2003). *Cancer Res***63**, 6178–6186
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. *TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity.* (2011). *Nature* **473**, 343-348.
- Wu, S.C. and Zhang, Y. *Active DNA demethylation: many roads lead to Rome.* (2010). *Nat Rev Mol Cell Biol***11**, 607-620.
- Zuchegna, C., Aceto, F., Bertoni, A., Romano, A., Perillo, B., Laccetti, P., Gottesman, M.E., Avvedimento, E.V., Porcellini, A. *Mechanism of retinoic acid-induced transcription: histone code, DNA oxidation and formation of chromatin loops.* (2014). *Nucleic Acids Res.* **42**, 11040-11055.
- Zhu J.K. *Active DNA demethylation mediated by DNA glycosylases.*(2009). *Annu Rev Genet.* **43**, 143-66.

## LIST OF PUBLICATIONS

This dissertation is based upon the following publications:

1. Morano A, Angrisano T, **Russo G**, Landi R, Pezone A, Bartollino S, Zuchegna C, Babbio F, Bonapace IM, Allen B, Muller MT, Chiariotti L, Gottesman ME, Porcellini A, Avvedimento EV. **“Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene”**. Nucleic Acids Res. 2014 Jan;42(2):804-21. doi: 10.1093/nar/gkt920. Epub 2013 Oct 16. PMID: 24137009 [PubMed - indexed for MEDLINE] PMCID: PMC3902918

## Curriculum Vitae

Dr. Giusi Russo  
Via Stefano Piro, 15  
80034 Marigliano(NA),Italy  
(0039) 3204164285  
russo.giusi84@gmail.com

**Date of Preparation:** March 31, 2015

### Personal Data

Name	Giusi Russo
Date of Birth	January 31, 1984
Birthplace	Avellino
Citizenship	Italy

### Training

---

03/2012 – 02/2015	<b>Department of Molecular Medicine and Biotechnology</b> <i>PhD student in Molecular Pathology and Physiopathology, 27°cycle</i>	<i>Naples, Italy</i>
07/2013 – 09/2013	<b>Institute for Cancer Genetics, Irving Cancer Research Center</b> , Columbia University Medical Center. <i>Visiting researcher</i>	<i>New York, NY</i>
03/2008 – 02/2012	<b>Department of Molecular and Cellular Biology and Pathology</b> <i>Internship</i>	<i>Naples, Italy</i>

### Education

---

09/2008 – 09/2010	<b>Degree in Medical Biotechnology, University of “Federico II”</b> <i>Score: 110 / 110</i>	<i>Naples, Italy</i>
09/2003 – 09/2008	<b>Degree in Biotechnological Sciences, University of “Federico II”</b>	<i>Naples, Italy</i>

### Honors

---

**Teaching Experience and Responsibilities**

6/2012 – present	Senior student advisor and trainer Department of Molecular Medicine and Biotechnology	Naples, Italy
------------------	---	---------------

**Publications**

- 1/2014 Morano A, Angrisano T, **Russo G**, Landi R, Pezone A, Bartollino S, Zuchegna C, Babbio F, Bonapace IM, Allen B, Muller MT, Chiariotti L, Gottesman ME, Porcellini A, Avvedimento EV. **"Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene"**. Nucleic Acids Res. 2014 Jan;42(2):804-21. doi: 10.1093/nar/gkt920. Epub 2013 Oct 16.
- 3/2015 **G. Russo**, R. Landi, A. Pezone, C. Zuchegna, M. E. Gottesman, A. Porcellini, E. V. Avvedimento. **"Chromatin modification induced by DNA damage and repair DNA methylation induced by DNA damage and repair causes stochastic variations of gene expression"**. (2015) in preparation.

**Abstract**

- 9/2012 A. Pezone, R. Landi, G. Russo, A. Porcellini, V.E. Avvedimento. **"Dna oxidation, transcription and methylation: the same molecular machine with different facets"**.  
*Conference: European Environmental Mutagen Society (EEMS). Poland*
- 10/2012 **G. Russo**, C. Zuchegna, A. Pezone, R. Landi, A. Porcellini, V. E. Avvedimento. **"Targeted DNA methylation induced by homology-directed repair in mammalian cells"**.  
*Meeting: Joint National Ph.D. (poster presentation) Rimini, Italy*
- 10/2012 R. Landi, A. Pezone, C. Zuchegna, **G. Russo**, A. Porcellini, V. E. Avvedimento. **"Methylation-BER cycles drive estrogen induced transcription"**.  
*Meeting: Joint National Ph.D. Rimini, Italy*
- 10/2012 C. Zuchegna, **G. Russo**, R. Landi, A. Pezone, V. E. Avvedimento, A. Porcellini. **"Mechanism of retinoic acid-induced transcription: epigenetic changes, DNA oxidation and chromatin loops"**.  
*Meeting: Joint National Ph.D. Rimini, Italy*

- 10/2012 A. Pezone, R. Landi, **G. Russo**, C. Zuchegna, A. Porcellini, V. E. Avvedimento. *Rimini, Italy*  
**“Local DNA oxidation and DNA methylation set chromatin loops and drive the transcription cycles induced by estrogen.**  
*Meeting: Joint National Ph.D.*
- 10/2013 **G. Russo**, R. Landi, A. Pezone, S. Agnese and E. V. Avvedimento. *Napoli, Italia*  
**DNA methylation is a "scar" induced by DNA damage and homologous repair. Transcription reshapes methylation on the repaired gene.**  
Retreat: Dipartimento di Medicina Molecolare e Biotecnologie Mediche.

# Targeted DNA methylation by homology-directed repair in mammalian cells. Transcription reshapes methylation on the repaired gene

Annalisa Morano<sup>1,2</sup>, Tiziana Angrisano<sup>1</sup>, Giusi Russo<sup>1</sup>, Rosaria Landi<sup>1</sup>, Antonio Pezone<sup>1</sup>, Silvia Bartollino<sup>3</sup>, Candida Zuchegna<sup>4</sup>, Federica Babbio<sup>5</sup>, Ian Marc Bonapace<sup>5</sup>, Brittany Allen<sup>6</sup>, Mark T. Muller<sup>6</sup>, Lorenzo Chiariotti<sup>1</sup>, Max E. Gottesman<sup>7,\*</sup>, Antonio Porcellini<sup>4,\*</sup> and Enrico V. Avvedimento<sup>1,\*</sup>

<sup>1</sup>Dipartimento di Medicina Molecolare e Biotecnologie mediche, Istituto di Endocrinologia ed Oncologia Sperimentale del C.N.R., Università Federico II, 80131 Napoli, Italy, <sup>2</sup>IRCCS CROB, Dipartimento di Oncologia Sperimentale, via Padre Pio, 1 85028 Rionero in Vulture, Italy, <sup>3</sup>Dipartimento di Medicina e di Scienze della Salute, Università del Molise, 86100 Campobasso, Italy, <sup>4</sup>Dipartimento di Biologia, Università Federico II, 80126 Napoli, Italy, <sup>5</sup>Dipartimento di Biologia Strutturale e Funzionale, Università dell'Insubria, Varese 21100, Italy, <sup>6</sup>Department of Molecular Biology and Microbiology and Biomolecular Science Center, University of Central Florida, 12722 Research Parkway, Orlando, FL 32826, USA and <sup>7</sup>Institute of Cancer Research, Departments of Microbiology and Biochemistry and Molecular Biophysics, Columbia University Medical Center, New York, NY 10032, USA

Received July 1, 2013; Revised August 30, 2013; Accepted September 19, 2013

## ABSTRACT

**We report that homology-directed repair of a DNA double-strand break within a single copy Green Fluorescent Protein (GFP) gene in HeLa cells alters the methylation pattern at the site of recombination. DNA methyl transferase (DNMT)1, DNMT3a and two proteins that regulate methylation, Np95 and GADD45A, are recruited to the site of repair and are responsible for selective methylation of the promoter-distal segment of the repaired DNA. The initial methylation pattern of the locus is modified in a transcription-dependent fashion during the 15–20 days following repair, at which time no further changes in the methylation pattern occur. The variation in DNA modification generates stable clones with wide ranges of GFP expression. Collectively, our data indicate that somatic DNA methylation follows homologous repair and is subjected to remodeling by local transcription in a discrete time window during and after the damage. We propose that DNA methylation of repaired genes represents a DNA damage code and is source of variation of gene expression.**

## INTRODUCTION

DNA methylation is a feature of higher eukaryote genomes. It is thought to help organize large segments of noncoding DNA in heterochromatin and to contribute to genome stability (1). DNA methylation is critical during development in plants and mammals. In somatic cells, patterns of methylated CpGs are transmitted to daughter cells with high fidelity (2,3). Aberrant methylation, both hyper- and hypo-methylation, has been found in cancer cells (4).

There are two patterns of DNA methylation: (i) Stable methylation, which is the basis of imprinting, is inherited in a sex-specific fashion and is invariant among individuals and cell types. Loss or modification of stable methylation results in significant phenotypic and genetic alterations. (ii) Unstable or metastable methylation, which is variable among individuals and cell types.

Despite numerous analyses of the methylation profiles of single chromosomes, the regulation of DNA methylation is largely unknown. Somatic DNA methylation is associated with gene silencing and heterochromatin formation and is neither sequence- nor cell-specific.

We are investigating the nature of somatic DNA methylation and its link to gene silencing during neoplastic progression (5,6). Since formation of DNA double-strand

\*To whom correspondence should be addressed. Tel: +39081679047; Fax: +39081679233; Email: antonio.porcellini@unina.it  
Correspondence may also be addressed to Max E. Gottesman. Tel: +1 212 305 6900; Fax: +1 212 305 1741; Email: meg8@columbia.edu  
Correspondence may also be addressed to Enrico V. Avvedimento. Tel: +390817463251; Fax: +390817463308; Email: avvedim@unina.it

breaks (DSBs) and activation of DNA damage checkpoints may precede genomic instability (7) and DNA methylation and gene instability appear to be linked in cancer (8), we speculated that DNA methylation was associated with DNA damage and repair.

We previously reported that homology-directed repair (HDR) modifies the methylation pattern of the repaired DNA (9). This was demonstrated using a system pioneered by Jasin (10,11), in which recombination between partial duplications of a chromosomal Green Fluorescent Protein (GFP) gene is initiated by a specific DSB in one copy. The unique DSB is generated by cleavage with the meganuclease I-SceI, which does not cleave the eukaryotic genome. The DSB is repeatedly formed and repaired, until the *I-SceI* site is lost by homologous or nonhomologous repair or depletion of I-SceI enzyme. Recombination products can be detected by direct analysis of the DNA flanking the DSB or by the appearance of functional GFP (9).

Two cell types are generated after recombination: clones expressing high levels of GFP and clones expressing low levels of GFP, referred to as H and L clones, respectively. Relative to the parental gene, the repaired GFP is hypomethylated in H clones and hypermethylated in L clones. The altered methylation pattern is largely restricted to a segment just 3' to the DSB. Hypermethylation of this tract significantly reduces transcription, although it is 2000 bp distant from the strong cytomegalovirus (CMV) promoter that drives GFP expression (9,12). The ratio between L and H clones is ~1–2 or 1–4, depending on the insertion site of the GFP reporter. These experiments were performed in mouse embryonic (ES) or human cancer (Hela) cells. HDR-induced methylation was dependent on DNA methyl transferase I (DNMT1). Furthermore, methylation induced by HDR was independent of the methylation status of the converting template (9). These data, taken together, argue for a cause–effect relationship between DNA damage–repair and DNA methylation.

The link between DNA damage, repair and de novo methylation has been confirmed by other studies (13–15). We also note that genome wide surveys show that imprinted sites are historical recombination hot spots, reinforcing our conclusion and that of other workers, that DNA methylation marks the site of DNA recombination (16,17).

We report here that methylation induced by HDR is influenced by recruitment of Np95 and GADD45a to the DSB and that DNMT3a is also active at the DSB. We also show that methylation is reduced by transcription of the repaired region.

## MATERIALS AND METHODS

### Cell culture, transfections and plasmids

HeLa cells lines were cultured at 37°C in 5% CO<sub>2</sub> in RPMI medium supplemented with 10% fetal bovine serum (Invitrogen), 1% penicillin-streptomycin, and 2 mM glutamine.

HeLa-pDRGFP cells were obtained by transfection of HeLa cells with the pDRGFP plasmid. Briefly:  $5 \times 10^6$

cells were seeded in a 100 mm dish and transfected with lipofectamine as recommended by the manufacturer (Invitrogen) with 2 µg of linearized pDRGFP plasmid and selected in the presence of puromycin (2 micrograms/ml). Four clones were isolated and expanded, the remaining clones were screened for single pDRGFP insertion by quantitative Polymerase Chain Reaction (qPCR) [supporting information in (9)] and pooled (~200 clones with a pDRGFP copy number ranging from 0.8 to 1.2 copies/genome). Clone 3 is the same clone 3 described in (9); clone 4 is a subclone of the clone 2 assayed also by Southern Blot (9).  $10^6$  puromycin-resistant cells were transiently transfected by electroporation with 2.5 µg of plasmid DNAs and/or small interfering RNA (siRNA) (200 nM) as indicated in the Figures. After transfection cells were seeded at  $3 \times 10^5$  cells per 60 mm dish, 24 h post-transfection, cells were treated and harvested as described in figures. Pools of clones were generated in three independent transfections and frozen in aliquots. Transient transfections with I-SceI were carried at different times of culture after the primary transfection. Transfection efficiency was measured by assaying β-galactosidase activity of an included pSVβGal vector (Promega). Normalization by fluorescent-activated cell sorter (FACS) was performed using antibodies to β-gal or pCMV-DsRed-Express (Clontech). pEGFP (Clontech) was used as GFP control vector. The structure of the pDRGFP and other plasmids are described in the supplementary data (Supplementary Methods and Supplementary Figure S12).

### Nucleic acid extraction and quantitative reverse

#### Transcription Polymerase Chain Reaction, qPCR and PCR

Total RNA was extracted using Triazol (Gibco/Invitrogen). Genomic DNA extraction was performed as described in (9). cDNA was synthesized in a 20 µl reaction volume containing 2 µg of total RNA, four units of Omniscript Reverse Transcriptase (Qiagen), and 1 µl random hexamer (20 ng/µl) (Invitrogen). mRNA was reverse-transcribed for 1 h at 37°C, and the reaction was heat inactivated for 10 min at 70°C. The products were stored at –20°C until use. Amplifications were performed in 20 µl reaction mixture containing 2 µl of synthesized cDNA product or 0.1 µg of genomic DNA, 2 µl of 10X PCR buffer, 1.5 mM MgCl<sub>2</sub>, 0.5 mM dNTP, 1.25 unit of Taq polymerase (Roche), and 0.2 µM of each primer on a TC3000G thermocycler (Bibby Scientific Italia). The number of cycles was selected and validated by running several control reactions and determining the linear range of the reaction. 15 µl of the PCR products were applied to a 1.2% agarose gel and visualized by ethidium bromide staining. Densitometric analysis was performed using a phosphorimager. Each point was determined in at least three independent reactions. Quantitative reverse Transcription Polymerase Chain Reaction (qRT-PCR) and qPCR were performed three times in six replicates on a 7500 Real Time-PCR on DNA template (RT-PCR) System (Applied Biosystems) using the SYBR Green-detection system (FS Universal SYBR Green MasterRox/Roche Applied Science). The complete list of oligonucleotides is reported in Supplementary Table S1.



### FACS analysis

HeLa-DRGFP cells were harvested and resuspended in 500  $\mu$ l of phosphate buffered saline (PBS) at density of  $10^6$  cells/ml. Cell viability was assessed by propidium iodide (PI) staining. Cytofluorimetric analysis was performed on a 9600 Cyan System (Dako Cytometrix) or FACScan Flow Cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). PI positive cells were excluded from the analysis by gating the PI-negative cells on a FSC-Linear versus FL2H-Log plot. GFP<sup>+</sup> cells were identified by using a gate (R1 in Supplementary Figure S3A) on a FL1H-Log versus FL2H-Log plot after sample compensation for FL1 versus FL2 channels. L and H cells were identified on FL1H Histogram of the R1-gated cells with two range-gate, as shown in Figure 1. The same gate was used for all cytofluorimetric determinations.

Cell cycle analysis was carried out by FACS:  $1 \times 10^6$  cells were resuspended in 1 ml of PBS and fixed 10 ml of ice-cold 70% ethanol. After 3 h, the cells were washed and stained for 30 min at room temperature with 0.1% Triton X100, 0.2 mg/ml Dnase-free RnaseA, 20  $\mu$ g/ml PI. Fluorescence was evaluated by FACS and analyzed by ModFit LT 2.0 (Verity Software House, Topsham, ME, USA).

Population comparison was performed using the Population Comparison module of the FlowJo software (Tree Star, Inc., Ashland, OR). Difference in fluorescence intensity (mean) was determined using the matched pairs Student's *t* test.

### Bisulfite DNA preparation, PCR and sequence analysis

Sodium bisulfite analysis was carried out on purified genomic DNA and on 'chromatinized' DNA. The full list of the buffer formulation is reported in the Supplementary Methods (Buffers Formulation). Chromatinized DNA was obtained as follows:  $10^7$  cells were fixed at 4°C temperature with 1% formaldehyde for 3 min. The reaction was stopped with glycine to a final concentration of 125 mM. Nuclei were isolated and permeabilized by incubating cells for 20 min in Buffer A, 20 min in Buffer B and then resuspended in Buffer C (see Buffers Formulation in Supplementary Methods). Nuclei or purified genomic DNA was heat denaturated (96°C for 10 min) incubated in a fresh solution containing 5 M sodium bisulfite and 20 mM hydroquinone and incubated at 37°C for 18 h. The cross-link was reversed, and proteins were digested with proteinase K (50  $\mu$ g/ml at 55°C for 2 h, and then at 65°C overnight). DNA was purified using a Wizard genomic purification kit (Promega), and then disulphonated by incubation for 15 min with NaOH to a final concentration of 0.3 M, neutralized with ammonium acetate to a final concentration of 3 M, and purified by ethanol precipitation. DNA was amplified by PCR using primers, listed in the Supplementary Table S1, using Taq polymerase, which is able to copy deoxyuridine, cloned in TOPO TA vector (Invitrogen), and sequenced with the M13 reverse primers.

### Chromatin Immunoprecipitation

Cells were transfected and/or treated as indicated in the legends of the figures. The cells ( $\sim 1 \times 10^6$ ) were fixed by

adding formaldehyde directly in the culture medium to a final concentration of 1% for 10 min at room temperature and washed twice using ice cold PBS containing  $1 \times$  protease inhibitor cocktail (Roche Applied Science) and 1 mM Phenylmethylsulfonyl Fluoride (PMSF). Fixed cells were harvested and the pellet was resuspended in 200  $\mu$ l of sodium dodecyl sulphate Lysis Buffer (ChIP Assay Kit/Upstate). After 10 min incubation on ice, the lysates were sonicated to shear DNA to 300- and 1000-bp fragments. Sonicated samples were centrifuged and supernatants diluted 10-fold in the ChIP Dilution Buffer (ChIP Assay Kit/Upstate). An aliquot (1/50) of sheared chromatin was further treated with proteinase K, phenol/chloroform extracted and precipitated to determine DNA concentration and shearing efficiency (input DNA). The chromatin immunoprecipitation (ChIP) reaction was set up according to the manufacturer's instructions. Briefly, the sheared chromatin was precleared for 2 h with 20  $\mu$ l of protein-A or protein-G agarose (Upstate) and 2  $\mu$ g of nonimmune IgG (New England Biolabs). Precleared chromatin was divided in two aliquots and incubated at 4°C for 16 h with 20  $\mu$ l of protein-A/G agarose and 2  $\mu$ g of the specific antibody (Np95, generated and characterized by IM Bonapace; RNA Pol II from Upstate cat. # 05-623; DNMT1, DNMT3a and DNMT3b from Abcam, cat. # ab-13537, ab-2850 and ab-2851, respectively) and nonimmune IgG respectively. Agarose beads were washed with wash buffers according to the manufacturer's instructions and immunoprecipitated DNA was recovered and subjected to qPCR using the primers indicated in the legend of the specific figures and in Supplementary Table S1.

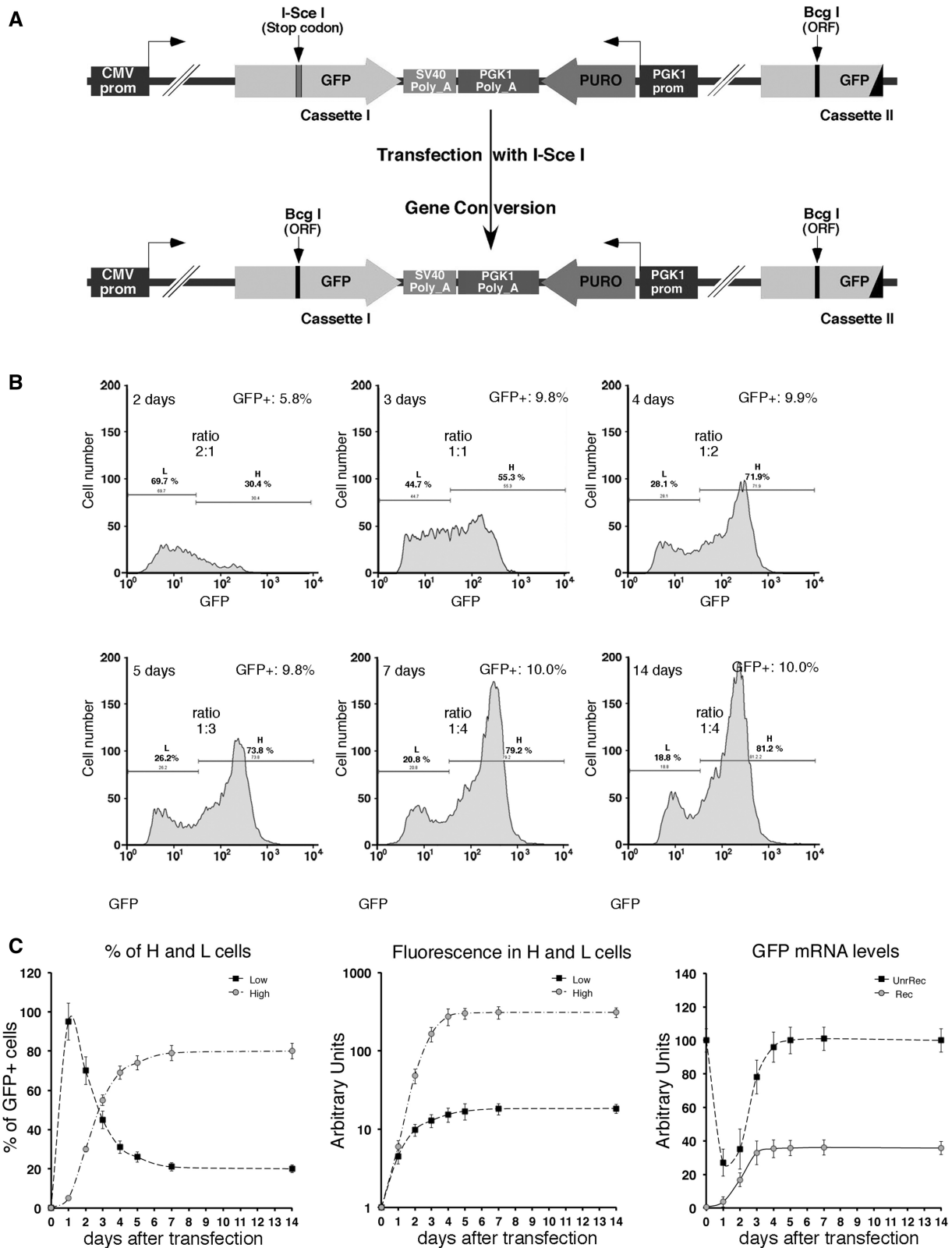
### Methylated DNA immunoprecipitation

Cells were transfected and/or treated as indicated in the legend of the figures. The cells ( $\sim 5 \times 10^6$ ) were harvested and genomic DNA extracted as described above. Ten micrograms of total genomic DNA were digested in 200  $\mu$ l for 16 h with restriction endonuclease mix containing 30 U each of Eco RI, Bam HI, Hind III, Xba I, Sal I (Roche Applied Science), phenol/chloroform extracted, ethanol precipitated and resuspended in 50  $\mu$ l of Tris-HCl/EDTA buffer (10 mM Tris-HCl pH 7.8, 1 mM EDTA) (TE) buffer. An aliquot (1/10) of digested DNA was used as input to determine the DNA concentration and digestion efficiency. Methylated DNA immunoprecipitation (MEDIP) was performed essentially as described (18) except that 2  $\mu$ g of antibody specific for 5mC (Abcam cat. # ab-124936) were used to precipitate methylated DNA from 5  $\mu$ g of total genomic DNA. H19 and UE2B were used to control in each experiment the efficiency of 5mC immunoprecipitation; the CpG island located to 5' end of human beta-actin was used as undamaged transcribed DNA gene control.

### Statistical analysis

All data are presented as mean  $\pm$  standard deviation in at least three experiments in triplicate ( $n \geq 9$ ). Statistical significance between groups was determined using Student's *t* test (matched pairs test or unmatched test were used as indicated in figure legends). Hierarchical clustering





**Figure 1.** HDR generates high and low GFP-expressing clones. (A) Structure of the integrated tester DRGFP plasmid before and after repair. The structure of the plasmid (10,11) has been verified by sequence analysis. The boxes and arrows with different grayscales represent the structural elements of the integrated nonrecombinant (upper) and recombinant (lower) plasmid. The conversion of the *I-SceI* to *BcgI* restriction site marks the gene conversion event driven by the copy of GFP gene located at the 3' end of DRGFP (cassette II). (B) Generation and accumulation of high (H) and low (L) expressor cells following homologous repair. Kinetics of L and H clones accumulation. Cells containing a single copy of DRGFP (clones 3 and 4, see 'Materials and Methods' section) or pool of clones (shown here), characterized as described in 'Materials and Methods' section, were transfected with *I-SceI* and subjected to FACS analysis at the times indicated. GFP positive (GFP<sup>+</sup>) cells were identified by the R1 gate

(continued)

(Ward's criterion) analysis was performed using the *JMP Statistical Discovery*<sup>TM</sup> software by SAS, Statistical Analysis Software. Sequence analysis and alignments were performed using MegAlign software (a module of the Lasergene Software Suite for sequence analysis by DNASTAR) for MacOSX.

## RESULTS

### Repair-induced methylation at the 3' end of a DSB

The system we use to study DNA methylation induced by damage and repair relies on a single-copy integrated plasmid (DRGFP), which contains two inactive versions of GFP. Introduction of a DSB in one copy of the gene (cassette I) by expressing the nuclease I-SceI, generates a functional GFP only in cells in which the second copy of GFP (cassette II) provides the template to repair the DSB (10,11) (Figure 1A). Homologous repair both in pools and single clones generates cells expressing low (L clones) or high levels (H clones) of GFP. These clones can be tracked by FACS analysis, using bivariate plots and gating strategies.

The integrated DRGFP undergoes several cycles of cutting and resealing until the *I-SceI* site is lost by nonhomologous end joining (NHEJ) or homology-dependent repair (HDR). We defined the time window of HDR by monitoring the appearance of recombinant GFP DNA in the population of cells transiently expressing I-SceI. We also measured the levels of I-SceI protein in transfected cells to estimate the period of enzymatic cleavage. Supplementary Figure S1A shows that the levels of recombinant GFP reached a plateau 3 days after transfection with I-SceI. The enzyme accumulated between 12 and 24 h and progressively disappeared 48–72 h after transfection. The estimated half-life of I-SceI protein was between 12 and 24 h (Supplementary Figure S1B).

Having established that the bulk of repair activity occurred in 3 days, we monitored the appearance and stabilization of L and H clones during and after HDR (9). Figure 1B shows the accumulation of L and H cells after exposure to I-SceI in a pool of HeLa clones as well as in single insertion clones carrying DRGFP inserts at different loci (see the legend of Figure 1B). Three days after I-SceI transfection, when HDR was almost complete, L and H cells accumulated in a 1:1 ratio (Figure 1B). We have used time-lapse microscopy to monitor GFP appearance during 30 h after I-SceI induction. The Supplementary Movie shows the I and II/III cycles (relative to GFP expression) during repair and the appearance of H and L cells from single repair events. In the I cycle, H and L cells are generated; in the II/III cycle (H-H and L-L), the

phenotypes are stably propagated. Eventually, the ratio L/H cells changes as a function of time, until day 7 when the L to H ratio stabilized at 1:4 (Figure 1B). No further change was detected after numerous subsequent passages, and no new GFP clones appeared (data not shown). Note that this shift to high GFP-expressing cells occurred after DSB repair, and therefore represents an inherited epigenetic process.

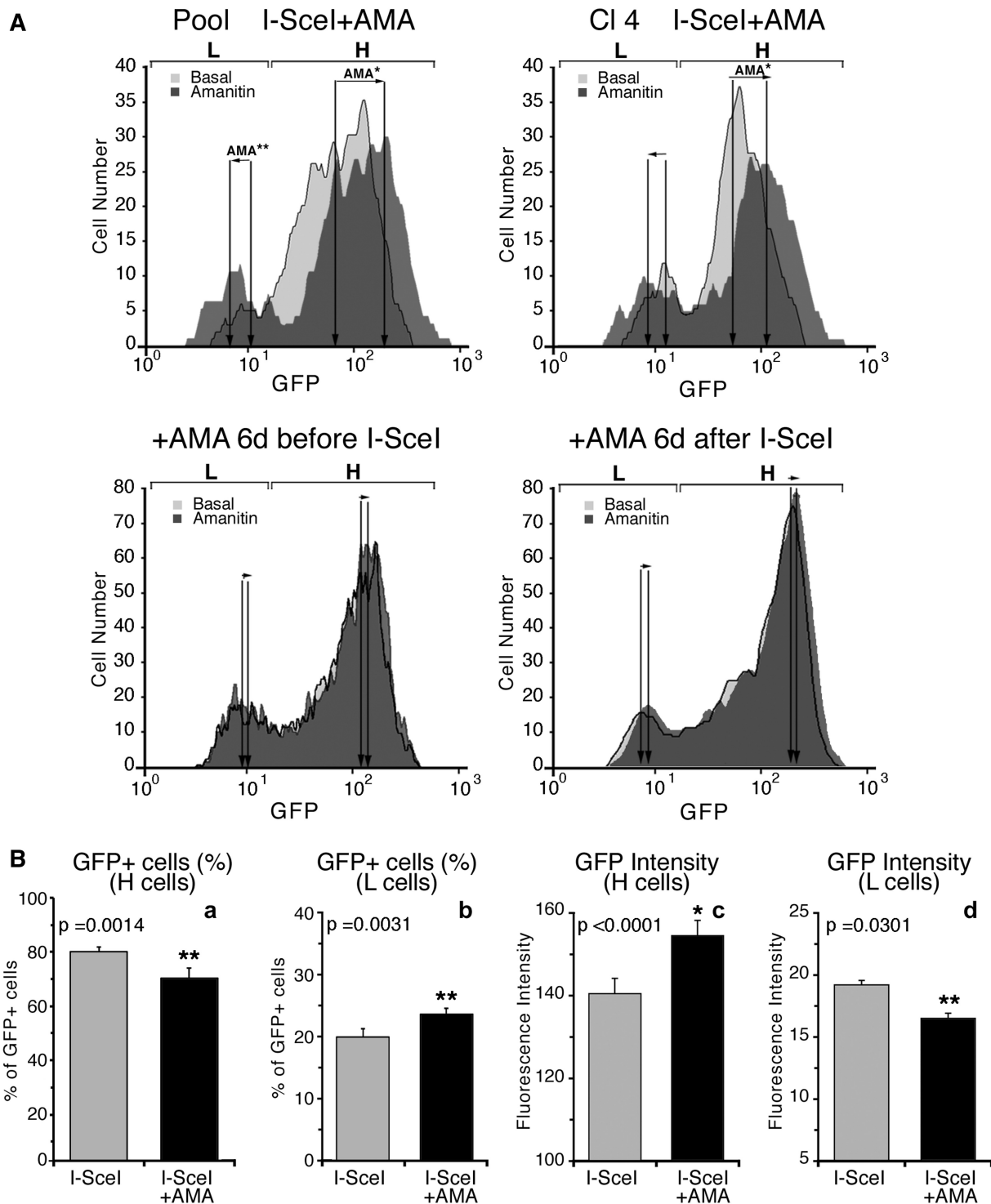
The drift toward H clones is detailed in Figure 1C. This figure also shows the levels of GFP mRNA as a function of time after transfection with the I-SceI plasmid. The changes in GFP mRNA concentrations correlate well with the fluorescence measurements that reflect GFP expression. We wondered if the time-dependent epigenetic changes were related to transcription of the GFP gene. This notion was tested by adding  $\alpha$ -amanitin during repair and following the appearance of L and H clones.  $\alpha$ -Amanitin inhibits translocation of elongating RNA polymerase II (Pol II) and increases the concentration of the polymerase on transcribed genes (19).

The pool of DRGFP clones, as well as one clone (Cl4), was transfected with I-SceI, and after 24 h, exposed to  $\alpha$ -amanitin for 24 h. Five days later (day 7 after I-SceI transfection), GFP expression was analyzed by cytofluorimetry. Exposure of the cells to the drug did not influence the rate of recombination (Supplementary Figure S2A). As expected, it significantly enriched GFP chromatin with Pol II molecules (Supplementary Figure S2B). Figure 2A and Supplementary Figure S3 show that  $\alpha$ -amanitin treatment of pooled cells (or clone 4) shifted the populations of L and H classes in opposite directions (see arrows AMA): L and H cells displayed on the average, lower or higher fluorescence intensity, respectively. Exposure to  $\alpha$ -amanitin 6 days before transfection with I-SceI or 6 days after did not affect the distribution of L and H clones (Figure 2A and Supplementary Figure S3A). Statistical analysis of the data of 28 independent experiments in which  $\alpha$ -amanitin was added during recombination to pools or single clones indicates that the results are reproducible (Figure 2B and Supplementary Figure S3B). Quantitative analysis of GFP fluorescence in H and L cells exposed to  $\alpha$ -amanitin during repair reveals that the fraction of L cells increased and that the GFP expression in these cells was markedly reduced. Conversely, the H cell fraction decreased, but the intensity of the fluorescent signal in these cells was enhanced (Figure 2B). We hypothesize that transient stalling of Pol II induced by  $\alpha$ -amanitin during repair, increased GFP methylation, yielding higher numbers of L clones and reducing the fraction of H clones.

We therefore asked if  $\alpha$ -amanitin altered the DNA methylation profile of the repaired GFP gene. Clones 3 and 4 were

### Figure 1. Continued

(see Supplementary Figure S3A) on a bivariate plot (FL1H versus FL2H) after I-SceI transfection. A representative experiment, displaying the L and H cells is shown. Each panel shows (i) the days after I-SceI transfection; (ii) total GFP positive cells (%); (iii) the range gates used to discriminate H and L cells; (iv) the ratio L/H, which reached a plateau 7–14 days after I-SceI transfection. Panel (C): the number (percent of total GFP<sup>+</sup> cells, left) and the fluorescence intensity (mean, center) of H and L cells derived from clones (not shown here) or pool of clones, based on at least five independent experiments. After 7–14 days, the L/H ratio and the intensity of L and H peaks stabilize. CMV-EGFP transfected cells, as control lines, display a single fluorescence peak (9). The right panel shows the relative levels, normalized to 18S RNA, of nonrecombinant (UnRec) and recombinant (Rec) GFP mRNA after I-SceI transfection (see 'Materials and Methods' section).



**Figure 2.** Synchronization of transcription by  $\alpha$ -amanitin during repair amplifies and consolidates L and H clones. (A) Cytofluorimetric analysis. Cells were exposed to  $\alpha$ -amanitin before, during or after I-SceI transfection as indicated on the top of each panel. A pool of HeLa DRGFP cells or a clone carrying a single insert were transfected with I-SceI expression vector, and 24h later, an aliquot was exposed for 24h to 2.5  $\mu$ M  $\alpha$ -amanitin. The cells were washed and cultured in normal medium for 5 days, when FACS analysis was carried out (day 7 after I-SceI transfection). The fluorescence plots of GFP positive cells (overlay of the histograms of RI gates, see Supplementary Figure S3) are shown. L and H represent the range gates to identify high and low expressors, respectively. The arrows, indicated by AMA, represent the shift of the mean fluorescence after  $\alpha$ -amanitin treatment. Differences between treatments were tested for statistical significance using Student's matched pairs *t* test: \**P* < 0.001, \*\**P* < 0.05. Under these conditions,  $\alpha$ -amanitin did not affect cell survival or growth rate. Five days after 24-h 2.5- $\mu$ M  $\alpha$ -amanitin treatment, transcription of several housekeeping genes was similar to untreated controls. The changes of GFP expression following the short treatment(s) with the drug during repair (24h after I-SceI transfection) were stable for up 3 months in culture. (B) Statistical analysis derived from 28 independent experiments, in which DRGFP cells were exposed to  $\alpha$ -amanitin during repair as indicated above. The panel shows the statistical significance of the means ( $\pm$  SD). Differences between treatments were tested for statistical significance using Student's matched pairs *t* test: \**P* < 0.001, \*\**P* < 0.05.

treated with  $\alpha$ -amanitin (6–24 h), sorted 5 days later into H and L clones and analyzed by MEDIP assay with specific antibodies against 5-methylcytidine (anti-5mC) with primers indicated in Figure 3A. Figure 3B shows that anti-5mC recognizes the region 3' to the *I-SceI* site in the repaired GFP. As predicted, the frequency of 5mC was higher in L clones than in H clones. Consistent with GFP expression profiles shown in Figure 2,  $\alpha$ -amanitin increased the levels of 5mC in the L clones. The changes in 5mC levels were specific to the recombinant GFP segment, since the methylation status of the  $\beta$ -actin 5' CpG island did not change (data not shown). Additionally, the methylation status of H19-DMR (Differentially Methylated Region), or UBE2B gene (NC\_000005.9), used as positive and negative controls of MEDIP immunoprecipitation, did not change after  $\alpha$ -amanitin (Figure 3C). To visualize directly the methylation status of the repaired segment of GFP in  $\alpha$ -amanitin-exposed cells, we performed bisulfite analysis of the GFP gene in treated cells (Supplementary Figure S4). The repaired GFP gene just 3' to the DSB was selectively hypermethylated or hypomethylated in L and H cells, respectively. Treatment with  $\alpha$ -amanitin for 6 or 24 h accentuated these alterations of methylation: L clones became more methylated and H clones less methylated than untreated cells. Longer exposure (48 h) to  $\alpha$ -amanitin did not significantly alter the methylation pattern seen at 6 or 24 h (see the legend of Supplementary Figure S4).

To explore further the local chromatin changes induced by methylation and the effects of  $\alpha$ -amanitin on this process, we analyzed sites on the GFP gene that were protected from bisulfite conversion. Briefly, chromatin of L and H cells was fixed with formaldehyde, heat denatured and exposed to bisulfite. By probing GFP DNA, we can detect specific DNA segments protected, most likely by bound proteins, that block C to T conversion by bisulfite or by structures preventing single-strand formation (Figure 3D). The protected segment of GFP corresponds to the region containing the methylated sites at the 3' end of *I-SceI*, as shown in Supplementary Figure S4. We found no protected areas in the H clones, whether or not they were treated with  $\alpha$ -amanitin. Exposure to  $\alpha$ -amanitin enhanced protection against bisulfite in most of the regions found to have increased DNA methylation after repair (compare Figures 3 and Supplementary Figure S4).

We propose that stalled RNA polymerase during repair favors the recruitment of enzymes that methylate the repaired DNA, consolidating the methylation of L clones. This event occurs only during repair because stalling Pol II before or after DSB repair does not modify GFP methylation and expression.

### Transcription modifies methylation of the repaired gene

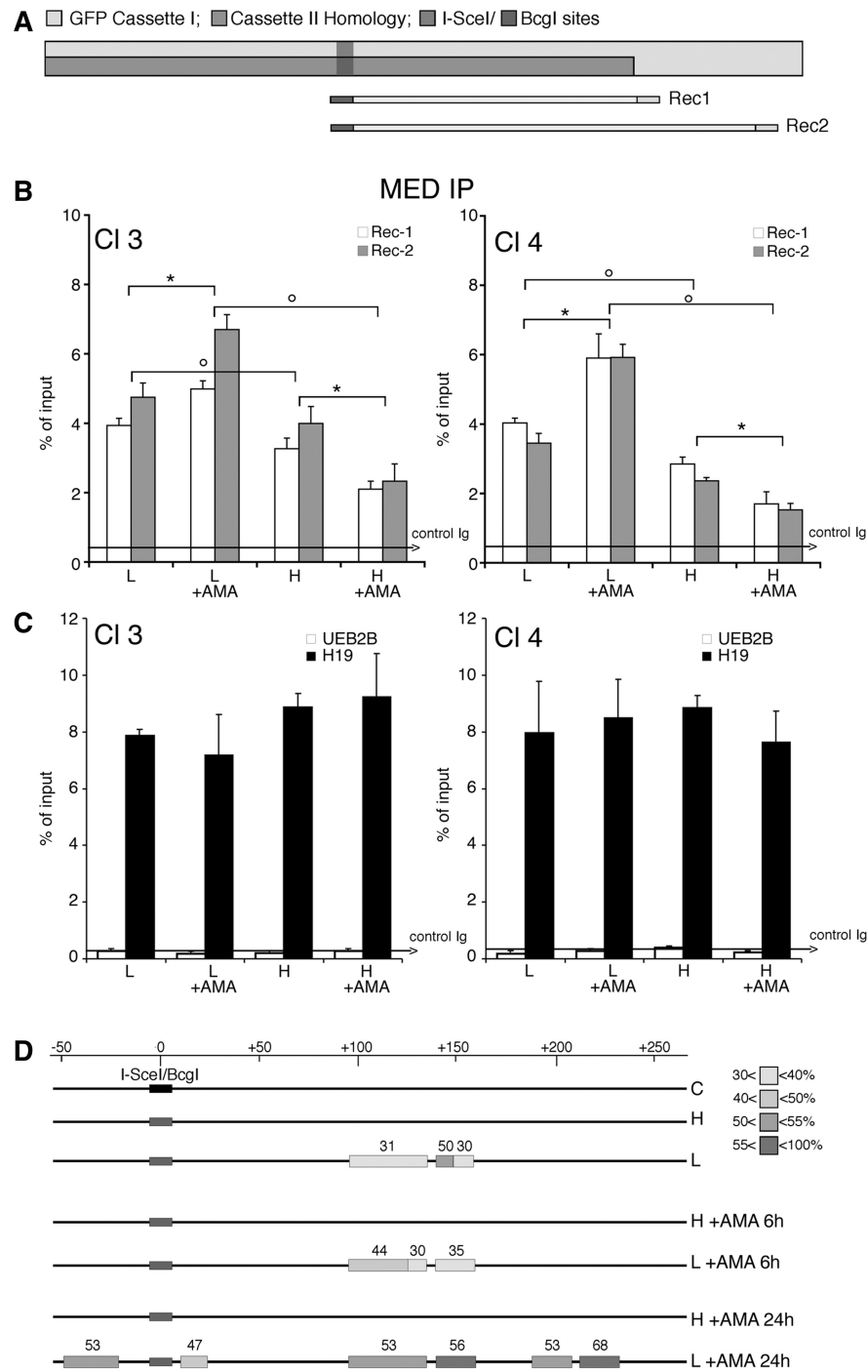
The  $\alpha$ -amanitin experiments suggest that the transcription machinery plays a major role in repair-induced methylation. We chose to inhibit transcription in a different fashion, by treating the cells with actinomycin-D (Act-D) for 6 h after repair. In contrast to  $\alpha$ -amanitin, Act-D depletes RNA polymerase II from chromatin (20).

We were unable to use Act-D during repair, owing to inhibition of HR by the drug (data not shown). After repair, 6 h exposure to Act-D did not alter DNA replication or HDR (legend of Figure 4). Under these conditions, the treatment with Act-D prevented the accumulation of H clones at 2 and 4 days later (5 and 7 days after *I-SceI* transfection), although the number of GFP<sup>+</sup> cells was similar in all samples (~9.5%), and the recombination frequency was unaltered (Figure 4B and data not shown). This finding suggests that the conversion of L to H cells after repair requires transcription (Figure 4B). To confirm the effectiveness of Act-D and to explore the mechanism of inhibition of H cell formation, we measured mRNA levels of several genes. Specifically, we analyzed the accumulation of stable and unstable RNAs: (i) recombinant (Rec) and nonrecombinant (UnRec) GFP; (ii) c-Myc (0.5–1 h half-life) (21); (iii)  $\beta$ -actin (8–12 h half-life; data not shown) (22); and (iv) 18S ribosomal RNA, 10 and 96 h after Act-D treatment. Figure 4C (left panel) shows the expected reduction in c-Myc, unRec and Rec mRNA levels 10 h after Act-D treatment (day 3). Rec mRNA was more stable than unRec mRNA. However, 96 h after Act-D exposure (day 7), UnRec and c-Myc mRNA concentrations returned to control values, whereas Rec mRNA levels remained lower than controls (Figure 4C, middle panel). Depletion of Pol II after Act-D exposure and the restoration of GFP-bound Pol II were confirmed by ChIP analysis of Un-Rec and Rec DNA (Figure 4C, rightmost panel). After 12–15 days, the increase of methylation and the inhibition of transcription of the GFP gene, induced by Act-D, progressively disappeared. Resumption of transcription promoted methylation loss during this period and accumulation of H cells from L cells (Figure 4B). These changes occurred only 2–3 weeks after the repair and were specific to repaired DNA because Act-D did not change the expression of undamaged GFP and, when administered 27–30 days after repair, did not modify GFP methylation (Supplementary Figure S5). We note that the time window of Act-D responsiveness (3–15 days after exposure to *I-SceI*) corresponds to the time required to stabilize the L/H cell ratio (Figure 1), suggesting that stabilization of the DNA–chromatin domain induced by HDR occurs in this interval. Collectively, these data indicate that after repair transcription converts a fraction of L to H cells by favoring loss of methylation.

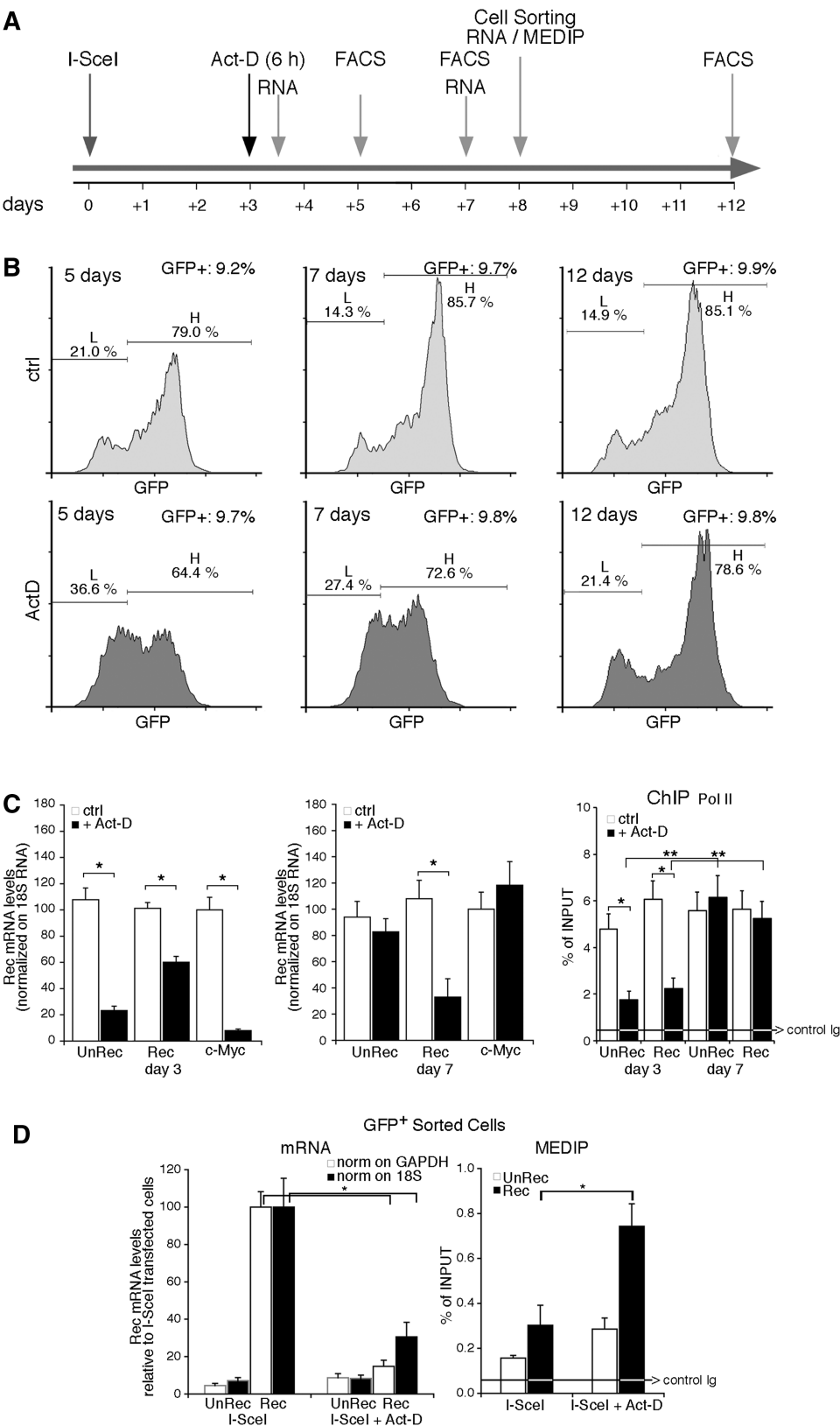
### Hierarchical clustering analysis of GFP methylation in repaired clones links discrete methylation states to gene expression variation

The data shown above indicate that the original methylation profiles induced by HDR are remodeled in a transcription-dependent fashion during the first 15 days after repair. The pattern eventually stabilizes, locking the epigenetic status of the repaired DNA in each cell (see Supplementary Movie, cycles I and II/III). By using hierarchical clustering analysis of bisulfite-treated GFP molecules before and after HDR, we were able to track and identify the original methylation profiles (epialleles) induced by HDR and modified during transcription. We





**Figure 3.** DNA methylation and chromatin modifications of the DSB region in cells exposed to  $\alpha$ -amanitin during repair. (A) Location of BclI, Rec1 and Rec2 primers, which recognize selectively recombinant GFP. Cassette I and II refer to Figure 1. (B) MEDIP with anti-5mC antibodies of recombinant GFP gene. Clones 3 and 4 were treated with  $\alpha$ -amanitin for 24 h as described in Figure 2 and sorted 5 days after I-SceI as described in 'Materials and Methods' section. Content of 5mC is higher in L cells compared with H cells,  $\alpha$ -amanitin also increases the levels of 5mC in L cells and lowers them in H cells. The results are similar for both amplicons (REC1 and REC2). All data derive from three independent experiments performed in triplicate (mean  $\pm$  SD;  $n = 9$ ). Differences between treatments were tested for statistical significance using Student's matched pairs  $t$  test:  $*P < 0.01$  as compared with the each control ( $\alpha$ -amanitin treated versus untreated cells). Differences between cells (H versus L) were tested for statistical significance using Student's  $t$  test:  $P < 0.01$ . (C) MEDIP analysis of the methylated H19 DMR (differentially methylated region) and the hypomethylated UBE2B genes in clone 3 and 4, treated with  $\alpha$ -amanitin, as indicated in B. Longer exposure (48 h) to  $\alpha$ -amanitin did not significantly alter the methylation pattern seen at 6 or 24 h assayed by bisulfite analysis (Supplementary Figure S4). (D) Bisulfite protection of GFP chromatin in L and H cells. Clone 4 cells were treated with  $\alpha$ -amanitin for 6 or 24 h after transfection and sorted as indicated in 'Materials and Methods' section. Chromatin was purified as described in 'Materials and Methods' section, denatured and treated with sodium bisulfite. DNA was extracted, amplified, cloned in TOPO TA vector and sequenced. The amplified segment corresponds to the Rec1 region and primers were designed for the bisulfite-converted (+) strand. The boxes represent stretches of nonconverted dCs present in the GFP sequence. At least 15 independent GFP molecules were analyzed for each treatment, including cells not exposed to I-SceI (C). The numbers with the grayscale boxes represent the percentage of the molecules protected from bisulfite conversion in the regions indicated by boxes. The scale shows the coordinates of the GFP sequence relative to the DSB (indicated as 0 or I-SceI/BclI site).



(continued)

also were able to link the methylation states of epialleles to GFP expression levels, since the bisulfite analysis was carried out on fluorescent-sorted cells. Clones expressing intermediate levels of GFP (L2 and H2) contain a set of GFP epialleles originating from a common GFP precursor segregating in the L fraction. This epiallele precursor in L cells generates many similar epialleles as a result of losing methyl groups (Supplementary Figures S6 and S7). These sites are shared by L2 and H2 clones and are located in 2 symmetric domains downstream of the DSB, spanning the length of a nucleosome (150 bp) (Supplementary Figure S7C and D). The sites are demethylated by 5-AzaC and methylated by Act-D treatments (Supplementary Figure S7C and D or data not shown). These data definitely link gene expression to specific methylation states and explain the stochastic expression of GFP after HDR (see Supplementary Movie).

### DNMT3a is transiently recruited to repaired GFP and stimulates DNA methylation

We previously reported that the hypermethylated L cell population was not found in a mutant lacking the maintenance DNMT1. In contrast, hypermethylation of the repaired gene was seen in both DNMT3a<sup>-/-</sup> and DNMT3b<sup>-/-</sup> mutants (9). However, loss of methylation induced by repair in stable DNMT1 mutant cells may be the indirect consequence of lack of propagation of methylation in daughter cells by DNMT1. Since large stretches of DNA are resynthesized during homologous recombination and are devoid of methylation marks, it is possible that de novo DNMTs such as DNMT3a and 3b have a role during or early after repair, and that DNMT1 may propagate the methylation marks set by DNMT3a and/or 3b during replication. To investigate this possibility, we analyzed the recruitment of DNMT3a and 3b to the I-SceI-cleaved chromatin. Figure 5A and B show that both DNMT3a and DNMT3b were recruited to the I-SceI site 24 h after the onset of DSB formation and rapidly disappeared (48 h). We then selectively silenced DNMT3a and 3b during repair and analyzed the

distribution of L and H cells. Figure 5C shows that the yield of L cells was significantly reduced and both the number and GFP fluorescence intensity of H cells increased when DNMT3a expression was silenced. In contrast, depletion of DNMT3b did not alter the ratio of L and H cells (Figure 5C). Expression of wild-type enzyme in DNMT3a-silenced cells prevented the loss of L cells. The changes of GFP expression levels were caused by DNA methylation, since the rescue of L cells by DNMT3a was prevented by treatment with 5azadC (Figure 5C).

In conclusion, we propose that DNMT3a helps the formation of hypermethylated clones and DNMT1 propagates these methylation patterns through at least several generations. This finding reinforces the notion that maintenance and de novo methyl transferases cooperate (23).

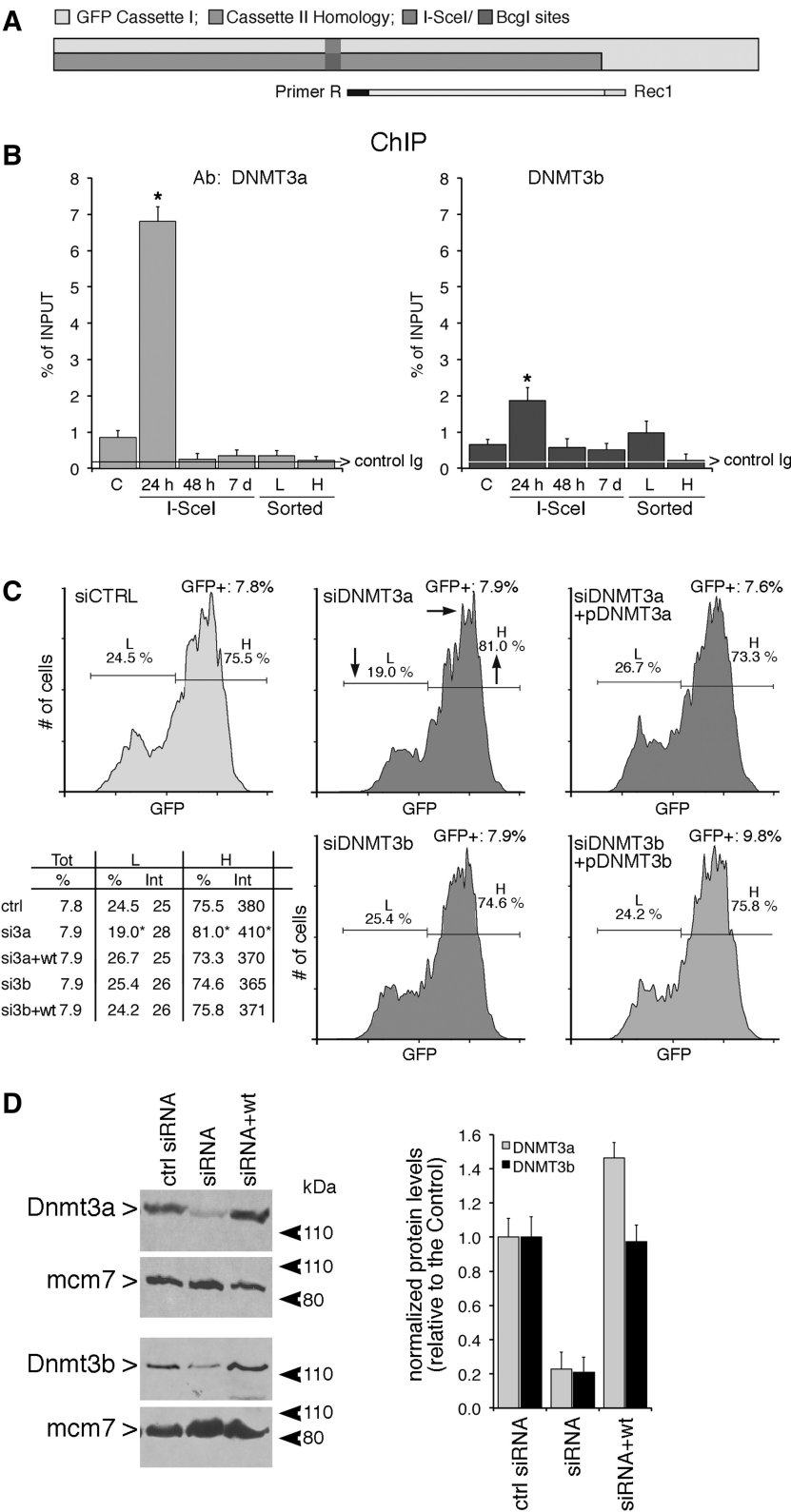
### Np95 is recruited to repaired GFP and stimulates DNA methylation

We reported that DNMT1 was required for hypermethylation of repaired GFP. We now ask if proteins that modify DNMT1 activity influence DNA methylation at the repaired DSB. We probed for Np95 (also known as UHRF1 or ICBP90), a protein that binds to DNMT1, DNMT3a, DNMT3b and PCNA and stimulates methylation of hemi-methylated DNA (24–26). ChIP analysis of GFP chromatin from clones 3 and 4 showed that Np95 preferentially accumulated on the repaired chromatin of the L clones. Treatment with  $\alpha$ -amanitin during repair significantly amplified or decreased Np95 recruitment to GFP chromatin in L or H cells, respectively (Figure 6A). Note that the binding of Np95 to H19, UEB2B or  $\beta$ -actin CpG island chromatin was unaffected by  $\alpha$ -amanitin (Figure 6B and data not shown). Thus, the association of Np95 with the DSB of GFP DNA appears to be linked to hypermethylation and reduced GFP expression in the L cell population.

To test whether Np95 recruitment to recombinant chromatin was relevant to repair-induced methylation, we selectively silenced Np95 expression during recombination.

#### Figure 4. Continued

(B) FACS analysis (a representative of five independent experiments) was performed as described in Figure 1 at 5, 7 and 12 days after I-SceI transfection (2, 4 and 9 days after Act-D treatment, respectively). Panel (C) Left. GFP mRNA accumulation assayed by qPCR after Act-D treatment (3 days after I-SceI transfection and 10 h after Act-D, or 7 days after I-SceI and 96 h after Act-D) normalized to 18S RNA. Recombinant (Rec) and nonrecombinant (UnRec) mRNA levels are expressed as percent of untreated levels  $\pm$  SD because the absolute mRNA levels cannot be compared because of the differences of the efficiency of the primers. The same results were obtained normalizing GFP RNA to GAPDH mRNA. Differences between treatments were tested for statistical significance using Student's matched pairs *t* test: \**P* < 0.01 as compared with the each untreated control. Right. RNA polymerase II recruitment on recombinant and nonrecombinant GFP chromatin after Act-D treatment. ChIP with anti-Pol II large fragment antibodies of chromatin extracted from Act-D-treated cells 3 days after I-SceI transfection (10 h after Act-D) or 7 days after I-SceI (96 h after Act-D). \**P* < 0.01 compared with the each untreated control; \*\**P* < 0.01, 3 days compared to 7 days time point; the average of immunoprecipitated DNA with a control Ig is reported on the bar graph. (D) GFP mRNA levels and MEDIP assay at day 8 on sorted GFP<sup>+</sup> cells. Left: Recombinant (Rec) and nonrecombinant (UnRec) primers were used to quantify GFP mRNA by qPCR and to measure the contamination of nonrecombinant GFP negative cells. The values were normalized to GAPDH (white columns) or 18S (black columns) RNAs. Rec mRNA levels are shown as percent of the levels found in control cells (I-SceI transfected/Act-D untreated cells); UnRec mRNA levels are expressed as percent of control (untransfected DRGFP cells) (mean of three experiments in triplicate  $\pm$  SD). \**P* < 0.01 as compared with untreated control. Right: 5mC content was carried out on sorted GFP<sup>+</sup> cells (H and L) as indicated in panel A. Specifically, we analyzed the 5mC content of (i) a segment of the GFP promoter, 1 kb upstream the DSB (oligo b and c, see Supplementary Table S1); (ii) the region 3' to the DSB, which was methylated by HDR; and (iii) H19 and UEB2B genes, as controls of hypermethylated and undermethylated genes, respectively, and to monitor the efficiency of MEDIP assays. The 5mC levels in these regions, except the segment 3' to the DSB, were not modified by 6 h Act-D treatment (data not shown). 5mC levels are expressed as percentage of input (mean  $\pm$  SD of three experiments in triplicate); the average of immunoprecipitated DNA with a control Ig is reported on the bar graph. \**P* < 0.01 as compared with the each untreated control. Act-D, administered 27, 30 and 35 days after I-SceI for 6 h, transiently inhibited transcription, but did not change GFP gene methylation.



**Figure 5.** DNMT3a and 3b are recruited to the DSB early during repair, but only DNMT3a is necessary for generation of L cells (**A** and **B**) Recruitment of DNMT3a, DNMT3b to the *I-SceI* chromatin. Cells were transfected with *I-SceI* and 24 h, 48 h or 7 days later, were fixed, collected, chromatin-extracted and subjected to ChIP analysis with specific anti-DNMT3a and DNMT3b antibodies. The specific primers used to amplify the GFP cassette I are indicated in (**A**). Data represent the fraction of immunoprecipitated DNA relative to the input chromatin-DNA present in the reactions (% of input; mean  $\pm$  SD;  $n \geq 9$ ); the average of immunoprecipitated DNA with a control Ig is reported on each bar graph. \* $P < 0.01$ , paired  $t$  test. (**C**) Silencing the expression of DNMT3a reduces L cells. Cells were electroporated with the siRNA targeting DNMT3a and DNMT3b (see ‘Materials and Methods’ section and protocol S1) and analyzed 7 days later, when L and H cells were clearly separated. On the bottom left

(continued)



We measured GFP expression, DNA methylation in the repaired segment and the frequency of recombination. Figures 6C (left panel) and Supplementary Figure S8 show that silencing of Np95 expression significantly enhanced fluorescence intensity in both the L and H cell fractions. Np95 depletion did not affect recombination frequency (Supplementary Figure S8A) but induced loss of methylation at the 3' end of the repaired GFP gene (Figure 6C, right panel). Under the same conditions, Np95 depletion did not modify the methylation status of  $\beta$ -actin CpG island, or stably methylated gene, H19 (DMR) (see the legend of Figure 6C). Overexpression of mouse wild-type Np95 reversed the effects of the silencing and reduced GFP expression (Supplementary Figure S8B).

Np95 interacts with several proteins involved in chromatin remodeling, specifically those that set repressive marks on histones, such as SUV39 and EZH2 (27,28). Indeed, 24 h after DSB induction, the I-SceI chromatin shows an accumulation of histone repressive (H3K9 m2-m3) and a reduction of positive H3K4 (m2 and m3) marks, respectively [(13) and data not shown]. To test if SUV39 and EZH2, which also interact with DNMT1 (27,28), play a role on DNA methylation induced by damage and repair, we silenced their expression during repair and determined the distribution of L and H cells. Knockdown of these proteins did not significantly modify the intensity of the GFP signal in either L or H cells (Supplementary Figure S9A). Although a modest decrease in GFP expression in SUV39-depleted cells was caused by inhibition of recombination (Supplementary Figure S9C), the levels of GFP methylation were not modified in cells in which SUV39 and EZH2 were silenced (Supplementary Figure S9D).

### GADD45a binds DSB and inhibits de novo methylation induced by HDR

To identify a DNMT1 partner that inhibits DNA methylation during repair and generates H cells, we monitored GADD45a (G45a) expression and localization after DSB formation. We recently found that GADD45A binds hemi-methylated DNA, inhibits DNMT1 *in vitro* and *in vivo* and reduces the fraction of L cells (18), suggesting that GADD45A promotes loss of methylation on the repaired DNA (29,30).

We first measured GADD45A mRNA levels in cells exposed to I-SceI or to the DNA-damaging agent, etoposide. GADD45A mRNA was induced by I-SceI and decreased to pre-induced levels 48 h after I-SceI transfection (Supplementary Figure S10). We next asked if GADD45A accumulated on DNA during HDR. ChIP analysis shows that GADD45A was recruited to GFP

chromatin 48 h after I-SceI expression, confirming a previous observation (18). Recruitment of GADD45A, as well as DNMT1 and Pol II, was further stimulated by  $\alpha$ -amanitin (Figure 7A and B). Note that DNMT1 accumulation on MGMT and p16, genes normally methylated in HeLa cells, was not stimulated by I-SceI expression or  $\alpha$ -amanitin (Figure 7A, lower panel).

We next tested the effects of silencing GADD45A on recombinant DNA methylation. Figure 6C shows that GADD45A knockdown (Supplementary Figure S11) inhibited GFP expression at 2 and 4 days after the damage. However, although reproducible, this effect, which was not noted previously (18), was transient; it was statistically significant at day 2 and progressively disappeared at 4 and 7 days after I-SceI expression (Figure 7C and Supplementary Figure S11 panels A and C). The consequences on GFP expression of GADD45A silencing at 2 days were reversed by co-transfection with a mouse GADD45A expression vector (Figure 7D, left panel). GADD45A silencing did not alter the frequency of recombination (Supplementary Figure S11D) but methylation of GFP was significantly stimulated, as shown by MEDIP analysis (Figure 7D, right panel). Under the same conditions, GADD45A depletion did not modify the methylation status of  $\beta$ -actin CpG island or of stably methylated genes, such as H19 (DMR) (Figure 7D).

The transient effects of GADD45A depletion on GFP expression may be dependent on the transient rise of the protein (18) and mRNA levels during damage and repair (Supplementary Figure S10). To address this issue, we overexpressed the wild-type protein, 2 days after I-SceI transfection, when endogenous protein levels were already low. Under these conditions, GADD45A stimulated GFP fluorescence intensity in H cells for longer periods (4–7 days after I-SceI), but at day 10 from the DSB, the effects disappeared (Supplementary Figure S11E and data not shown). However, 1 month after the DSB or in cells expressing CMV-EGFP, forced expression or induction of GADD45A by etoposide did not modify GFP levels (see the legend of Figure 7).

Taken together, these results indicate that Np95 and GADD45A favor the generation of L and H cells, respectively, during HDR.

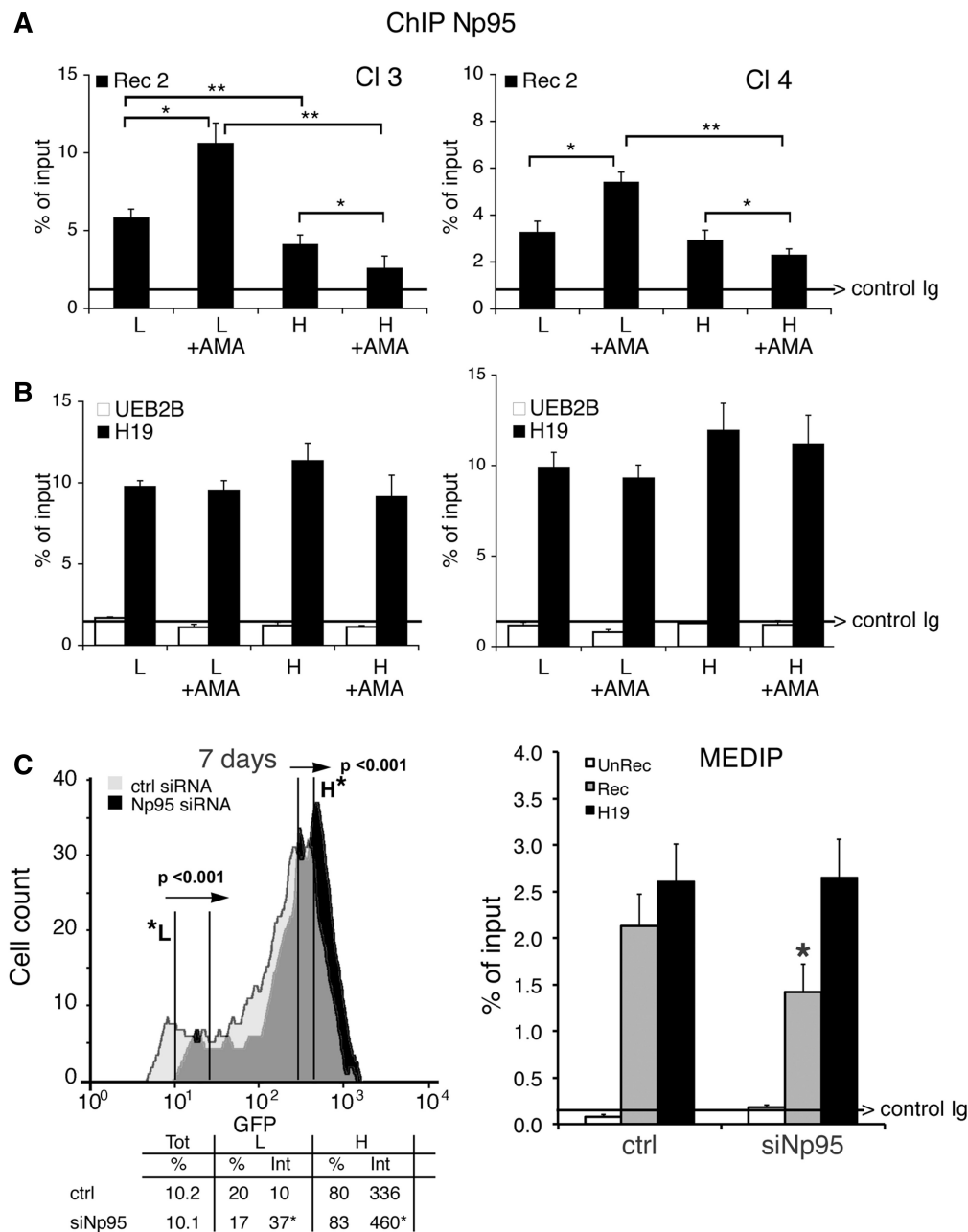
## DISCUSSION

### Mechanism of DNA repair-induced methylation

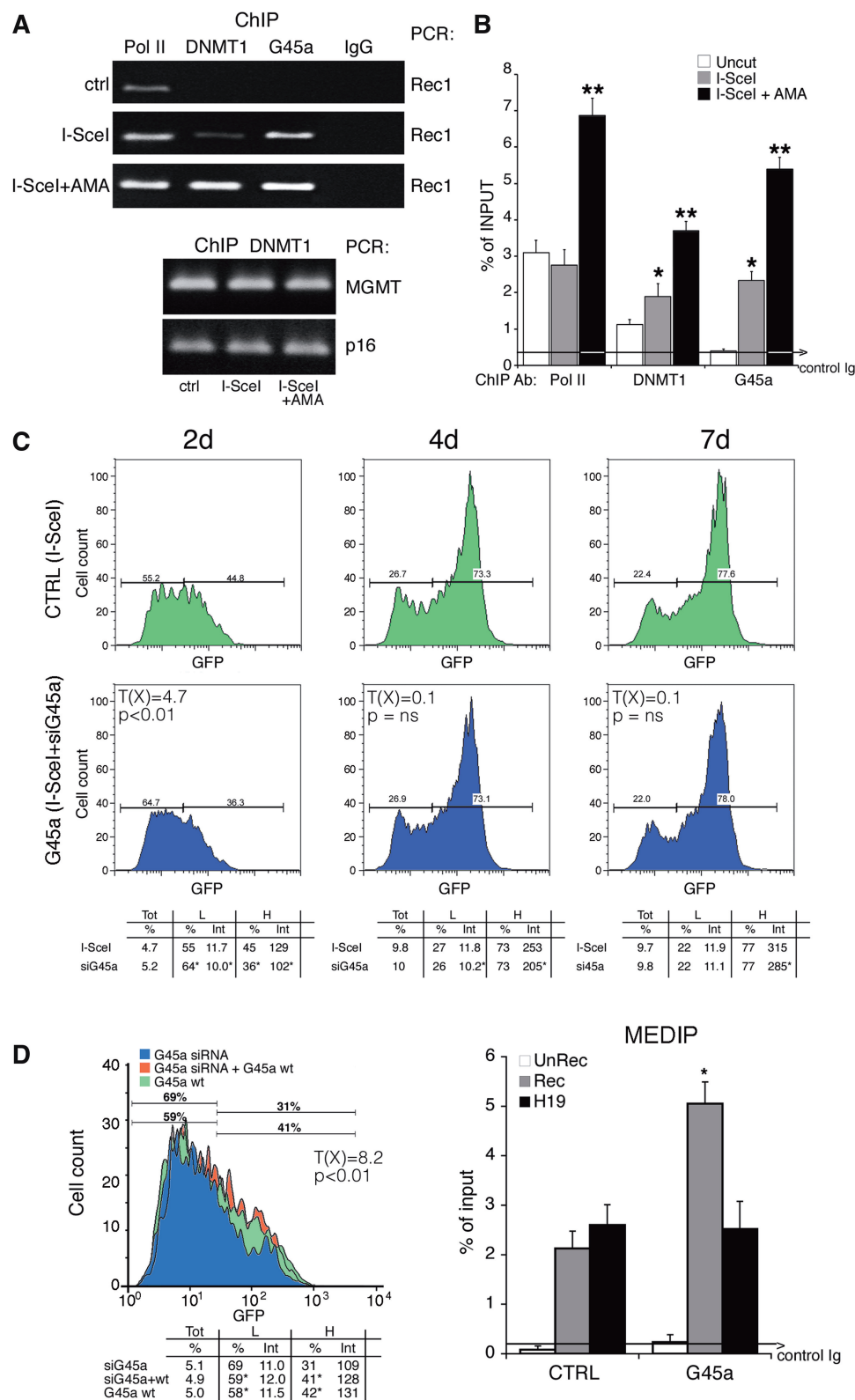
The results shown here argue for a link between HDR and DNA methylation at the site of a repaired DSB. Without DNA damage and repair, the expression of GFP is stable and uniform (Supplementary Figure S6, the red peak). DSB formation within GFP and repair by HDR

**Figure 5.** Continued

panel, statistical analysis derived from three independent experiments is shown. \* $P < 0.01$ , paired *t*-test comparing GFP intensity, Chi Square ( $\chi^2$ ) comparing the percentage of L/H cells. The horizontal and vertical arrows in the central inset indicate the shift in fluorescence intensity and in the distribution of L and H cells, respectively. Treatment with 5azadC (10  $\mu$ M for 2 days, 48 h after I-SceI transfection) rescued completely the loss of L cells (intensity and % GFP<sup>+</sup> cells) induced by DNMT3a overexpression in siDNMT3a-silenced cells (data not shown). (D) Western blot analysis of DNMT3a and 3b in silenced cells. Total cell extracts were prepared 48 h after electroporation and analyzed by immunoblot with the specific antibodies indicated. On the right is shown quantitative analysis derived from three immunoblots (mean  $\pm$  SD).



**Figure 6.** Np95 (UHRF1) is recruited to repaired GFP and stimulates DNA methylation. (A) ChIP with anti-Np95 antibodies of sorted cells exposed to  $\alpha$ -amanitin during repair. Clones 3 and 4 were transfected with I-SceI and treated with  $\alpha$ -amanitin for 24 h as described in Figure 2. The cells were sorted 5 days after I-SceI transfection and chromatin was collected from formaldehyde-fixed cells and subjected to ChIP analysis with specific antibodies to Np95. Primers Bcg and Rec2 were used to amplify recombinant GFP DNA. The data derive from three independent experiments performed in triplicate (mean  $\pm$  SD;  $n = 9$ ). Differences between treatments were tested for statistical significance using Student's matched pair  $t$ -test: \* $P < 0.01$  as compared with the each control ( $\alpha$ -amanitin treated versus untreated cells). Differences between cells (H versus L) were tested for statistical significance using Student's  $t$ -test: \*\* $P < 0.01$ . (B) ChIP analysis of Np95 on H19 DMR and UBE2B genes. qPCR was carried out with specific H19 DMR and UBE2B primers on the same samples indicated above. The fraction of immunoprecipitated DNA by control Ig is reported on each bar graph. (C) DRGFP cells (pool of clones; clone 3 and 4 are not shown here) were transiently transfected with a mixture of siRNAs targeting specifically human NP95 or control scrambled siRNA (ctrl) and the mouse I-SceI expression vector (see 'Materials and Methods' section). Six days later, the cells were subjected to FACS analysis and MEDIP. The left panel shows a representative experiment: arrows indicate the shift in silenced cells of GFP fluorescence intensity. The columns below the fluorescence plot show (i) the number of GFP<sup>+</sup> cells (Tot, expressed as percentage of cells); (ii) the mean fluorescence intensity (Int.); and (iii) Percentage of L and H cells on GFP<sup>+</sup> cells. Mean fluorescence intensity at day 7 increased from 10 to 37 in L cells and from 336 to 460 in H cells (left panel). FACS analysis was performed in triplicate in at least three experiments. Differences between treatments were tested for statistical significance using Student's matched pair  $t$ -test: \* $P < 0.001$  as compared with the each control (siRNA-treated versus untreated cells). Samples expressing NP95 wild-type and control cells were treated with 1  $\mu$ M 5azadC for 1 day (48 h after I-SceI), and the differences in fluorescence intensity was used to quantify methylation-dependent changes of GFP expression. The panel on the right shows the results of MEDIP immunoprecipitation with anti-5mC antibodies in control and siRNA-treated samples. Np95 depletion by siRNA did not modify the methylation status of stably methylated genes, such as H19 (DMR) and  $\beta$ -actin CpG island. \* $P < 0.01$  for  $t$ -value (matched pair test) relative to the cells treated with control scramble siRNA (CTRL). Data are expressed as mean  $\pm$  SD,  $n = 9$ ; the average of immunoprecipitated DNA with a control Ig is reported on the bar graph.



**Figure 7.** GADD45 is recruited to the DSB and transiently inhibits de novo methylation induced by HDR. (A) ChIP analysis with anti-GADD45A, DNMT1 and RNA polymerase II large fragment antibodies in HeLa cells, transfected (36 h) with I-SceI. Twelve hours after transfection, an aliquot of cells was treated for 24 h with  $\alpha$ -amanitin and processed as described in 'Materials and Methods' section. Bcg and Rec1 primers were used for semiquantitative PCR. Two methylated genes, MGM and p16, were used as controls for DNMT1 ChIP. Control IgG represents an average of nonimmune immunoglobulins used in ChIP. (B) Quantitative analysis by qPCR of at least three ChIP experiments in triplicate ( $n \geq 9$ ). Differences between treatments were tested for statistical significance using Student's matched pairs  $t$ -test: \* $P < 0.01$  as compared with uncleaved control; \*\* $P < 0.01$  compared with I-SceI. The average of immunoprecipitated DNA with a nonimmune Ig is reported on the bar graph. (C) DRGFP cells (pool of clones) were transiently transfected with siRNA pools targeting specifically GADD45A or control scrambled siRNA (ctrl) and

(continued)

significantly alter the methylation pattern of GFP in two steps. We propose that some actors at this phase are DNMT1/3a, Np95 and GADD45A, which transiently maintain the processed DSB 3' segment hemi-methylated, until replication generates methylated and hypomethylated daughter molecules. Figure 8 shows a simplified scheme describing the main events during and after DSB repair: (i) DNMT1 and DNMT3a are recruited to the DSB with Ga45a and NP95. DNMT3a is recruited in the first 24 h after damage and transiently cooperates with DNMT1 to methylate repaired DNA. At 48 h, Np95 and Ga45a amplify or limit transiently, respectively, DNMT1 activity on the hemi-methylated DNA, until replication duplicates the methylated and unmethylated DNA strands. This is better shown in the video presented in the Supplementary Movie, in which time lapse microscopy offers a unique snapshot into homologous repair. The appearance of the GFP signal in I-SceI synchronized cells can be monitored in the first and second cycle after recombination, relative to the GFP signal, generated by HDR. In the first cycle, H and L cells are formed from the same cell (square in the Supplementary Movie); in subsequent cycles, H and L cells stably propagate in culture the H or L phenotype (circle in the Supplementary Movie); (ii) After repair, transcription resumes at day 2–3 after DSB and progressively modifies local methylation profiles until the local domains of the *I-SceI* chromatin (loop A in H cells and loop C in L cells) are stabilize. We believe that this strand-selection mechanism accounts for the ~1:1 L/H ratio early after repair (Figures 1 and 4). In fact, GADD45A exerts its action early during repair (2–4 days), when the L/H ratio is close to 1 and before significant remodeling of methylation occurs (Figures 1, 4 and 7). Stalled RNA Pol II by  $\alpha$ -amanitin during repair may facilitate targeting DNMT1/3a complex to the 3' end (–) transcribed strand, thus promoting hyper-methylation of the 300 bp repaired DNA segment that lies 3' to the DSB relative to transcription orientation (Figure 3D and Supplementary Figure S4). The 3' end (+) strand, free from transcription proteins, probably is more prone to invade and find the homologous region to direct the annealing, the synthesis and ultimately the repair of the DSB (Synthesis Directed Strand Annealing, SDSA) (31). This

mechanism may account for the relatively high efficient HDR in our system.

### Remodeling of methylation by transcription after repair

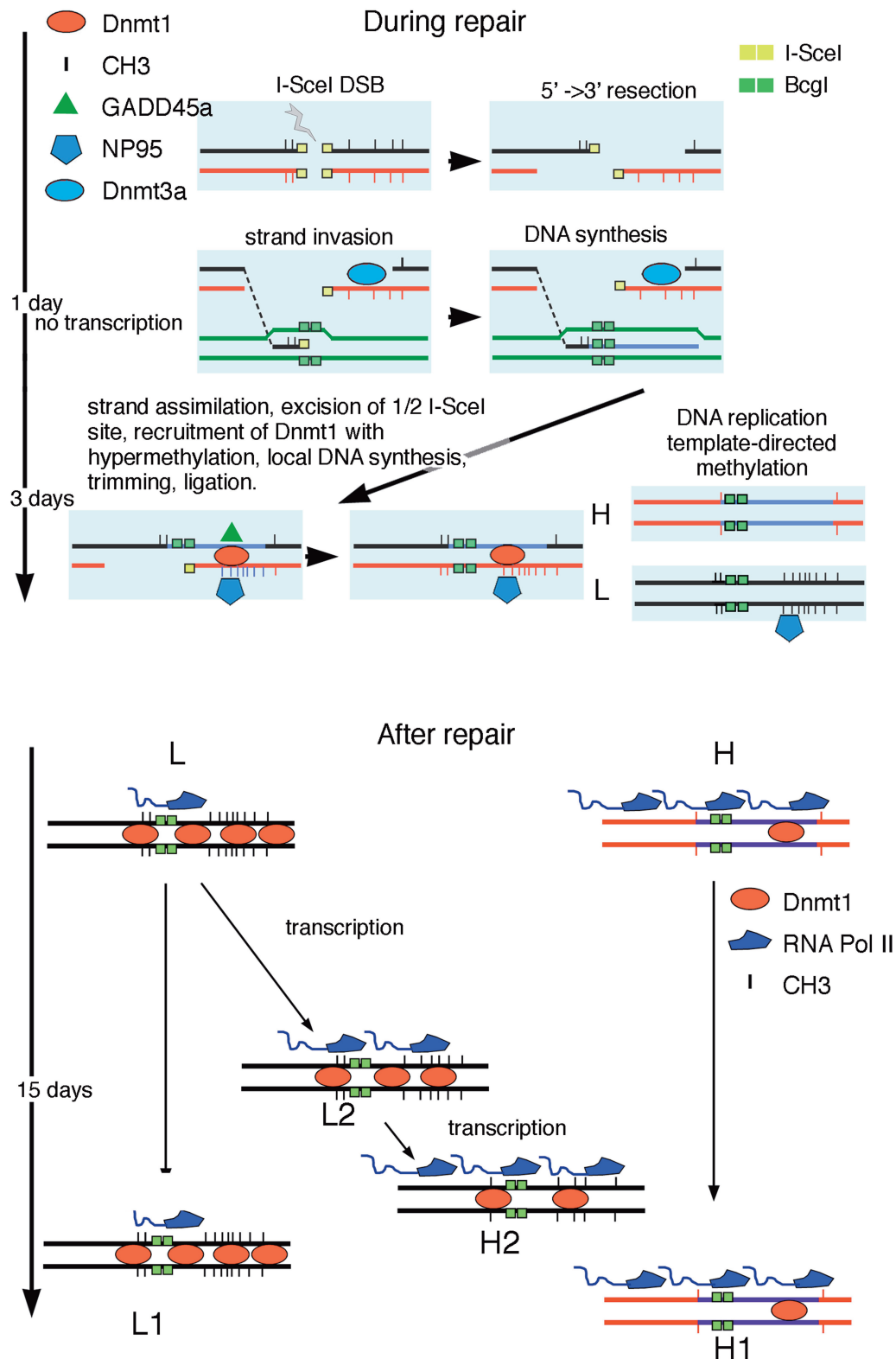
The second step of methylation induced by HDR begins ~48 h after formation of the DSB. At this point, repair is terminated, but chromatin and DNA continue to undergo epigenetic changes (9,13). H cells progressively increase and are similar in terms of methylation profile to a subpopulation of L cells (L2 in Supplementary Figure S6). Accumulation of these L2/H cells is favored by continuous transcription of GFP because transient inhibition of transcription after repair shifts the L/H ratio and favors accumulation of methylated clones (L2 in Figure 4D). We obtained essentially the same results shown in Figure 4 by transiently blocking transcription after repair with a dominant negative cdk9-expression vector, which inhibits phosphorylation of elongating RNA polymerase II (G.R., unpublished observations). However, 27, 30 and 35 days after DNA damage, inhibition of transcription by exposure to Act-D or expression of the dominant negative cdk9 did not alter methylation or expression of GFP (Supplementary Figure S5). These data indicate that inhibition of transcription per se does not trigger de novo methylation (33–35) and suggest that transcription may favor active demethylation. In fact, depletion of base excision repair (BER) enzymes (OGG1; APE1) or TDG increased methylation of repaired GFP similarly to Act-D treatment (data not shown), in agreement with the notion that transcription is associated with DNA methylation-demethylation (32,36) and DNA oxidation cycles (37). We note that the different effects of  $\alpha$ -amanitin and Act-D are related to the ability of these drugs to increase ( $\alpha$ -amanitin, Supplementary Figure S2B) or deplete (Figure 4D) RNA polymerase II from chromatin: (i) Stalled pol II during repair increases targeting and recruitment of DNMT1-Np95 on the DSB and favors accumulation of L clones; (ii) depletion or loss of pol II by slow resolution of Act-D/DNA inhibit transcription and active demethylation.

We suggest that transcription of damaged-repaired DNA is associated with stochastic replacement of methylated C by BER or nucleotide excision repair

### Figure 7. Continued

the I-SceI expression vector (see 'Materials and Methods' section). After 2, 4 and 7 days, the cells were subjected to FACS analysis as described in Figure 1. FACS analysis was performed in triplicate in at least three experiments. Differences in GFP expression between control and GADD45A-silenced cells were tested for statistical significance using the Chi Square test, T(X) (Population Comparison module of the FlowJo software from Tree Star). Differences of L and H (percentage and intensity) were tested for statistical significance using Student's *t* test: \**P* < 0.01 (see Supplementary Figure S10).  $\chi^2$  value (4.7), 2 days after I-SceI (control and GADD45A-silenced cells) (*P* < 0.01); at day 4 and 7,  $\chi^2$  value was not discriminant as day 2, although differences in fluorescence intensity of L and H cells between the control and GADD45A-silenced cells were significant (*P* < 0.02). All samples were treated with 1  $\mu$ M 5azadC for 1 day (48 h after I-SceI) to quantify methylation-dependent changes. (D) Left panel. Forced expression of GADD45A increases GFP expression. Cells were exposed to siRNA targeting the 3' UTR human GADD45A alone or in combination with vector expressing GADD45A. GFP fluorescence and Rec mRNA were analyzed 4 days later. The levels of specific GADD45A mRNA, the frequency of recombination in GADD45A-depleted cells and the statistical analysis of GFP expression are shown in Supplementary Figure S8. Differences between populations (control and GADD45A-silenced cells) were tested for statistical significance using the Chi Square test (Population Comparison module of the FlowJo software). Cells expressing CMV-EGFP-treated with etoposide or transfected with Ga45a expressing vector did not change GFP expression. Right panel. 5mC content of recombinant GFP in cells silenced for GADD45A. Four days after transfection, the cells were subjected to MEDIP assay. GADD45A depletion by siRNA did not modify the methylation status of stably methylated genes, such as H19 (DMR) and  $\beta$ -actin CpG island. \**P* < 0.01 for *t* value (matched pair test) relative to cells treated with control scramble siRNA (CTRL). All the samples in independent experiments were treated with 1  $\mu$ M 5azadC for 1 day (48 h after I-SceI) to quantify methylation-dependent changes. The average of immunoprecipitated DNA with a control Ig is reported.





**Figure 8.** Targeted methylation during and after homologous repair. The cartoon represents a schematic model illustrating the events during and after repair. The DSB undergoes 5' → 3' end resection and one of the 3' free single strand end invades the DNA of the GFP cassette II. The half *I-SceI* site is removed (flap removal) and new DNA is synthesized. Eventually, the invading strand returns to the original configuration and directs the synthesis of new DNA at the *Bcgl* site corresponding to the DSB, according to the SDSA model (Synthesis Directed Strand Annealing) (31). We propose that the asymmetric distribution of methylated CpGs in repaired GFP is caused by selective invasion of the (+) strand. The (−) strand, blocked by stalled RNA Pol II (DNMT1 and 3a), becomes a preferential target of DNMT1-Np95. The hemi-methylated DNA is replicated and generates H and L cells. After repair, transcription resumes and RNA Pol II-DNMT1 is associated with methylation/demethylation cycles (32) that in 15 days may remove some methyl groups in a subpopulation of L cells, leading to the conversion of L2 to H2 cells.

(NER) followed by repair synthesis (38,39). The events in this phase are distinct from those leading to the generation of H and L cells during repair, which are amplified by stalled RNA polymerase II and are dependent on Np95 and Ga45a. Under our conditions, GADD45A, transiently induced by DSB, recruited to the DSB, enhanced accumulation of hypomethylated clones (H) by inhibiting DNMT1 [Figure 7, Supplementary Figure S11 and (18)] and disappeared in 3–4 days. This inhibition may represent a barrier to spreading of repair-induced methylation. The opposing role of Np95 and GADD45A on DNMT1 activity is not new because DNMT1 stimulation and inhibition by Np95 and Ga45a, respectively, are required to maintain progenitor function in self-renewing somatic tissue (40).

### Evolution and stability of epialleles: qualitative analysis of methylation

Our data show that the repaired DSB in the GFP gene is marked locally by de novo methylation. Unlike the GFP system, in which we induced a site-specific DSB, DSBs in genomic DNA are essentially random in terms of sequence specificity, although the overall distribution is nonrandom, due to chromatin organization (41). Assuming that methylation marks these DSBs after homologous repair, the overall distribution of methylated sites in genomic DNA will appear random in the absence of selective pressure. We have extended our analysis to homologous targeting of GFP in ES cells and we find that genetically identical clones express variable GFP levels, due to de novo methylation or targeted gene (data not shown).

In our system, qualitative analysis of the methylation profiles, i.e. the location of methylated CpG in the various GFP molecules, 3' end to the DSB, is able to distinguish repaired GFP molecules from nonrecombinant or uncleaved molecules (Supplementary Figures S6 and S7). This discrimination is based on the relatedness of methylation profiles, not on the total methyl CpG content. GFP DNA molecules, shown in Supplementary Figures S4 and S6, can be considered epigenetic alleles because their methylation profiles are stable and are inherited in human and mouse cells over several generations. We have applied the same type of analysis shown in Supplementary Figures S6 and S7 to several somatically methylated genes and we find that the epialleles are stable, evolve rapidly following DNA damage and can be individually tracked in a complex mixtures of cells. HDR-induced specific methylation states may be ultimately responsible for stochastic gene expression in populations of mammalian cells.

In conclusion, we propose that DNA methylation represents a damage-repair code that modifies the expression of genes in cell populations and drives adaptation to environmental challenges. Selection of methylated alleles in each cell may be relevant for the rapid evolution of cancer cell phenotypes.

### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [42].

### ACKNOWLEDGMENTS

The authors thank A. Simeone and D. Acampora for providing us the ES clones with EGFP gene targeted to the *rosa 26* locus; G. Tell, G. Scholes and G. Diez Roux for helpful comments and F. Porcellini for video editing.

### FUNDING

AIRC IG [11364 to V.E.A.]; Epigenomics Flagship Project—EPIGEN (to C.N.R.) and Fondazione Medicina Molecolare e Terapia Cellulare, Università Politecnica delle Marche. Funding for open access charge: Epigenomics Flagship Project—EPIGEN (to C.N.R.).

*Conflict of interest statement.* None declared.

### REFERENCES

- Rizwana,R. and Hahn,P. (1999) CpG methylation reduces genomic instability. *J. Cell Sci.*, **112**, 4513–4519.
- Fu,A.Q., Genereux,D.P., Stöger,R., Laird,C.D. and Stephens,M. (2010) Statistical inference of transmission fidelity of DNA methylation patterns over somatic cell divisions in mammals. *Ann. Appl. Stat.*, **4**, 871–892.
- Probst,A.V., Dunleavy,E. and Almouzni,G. (2009) Epigenetic inheritance during the cell cycle. *Nat. Rev. Mol. Cell Biol.*, **10**, 192–206.
- Cho,Y.H., Yazici,H., Wu,H.C., Terry,M.B., Gonzalez,K., Qu,M., Dalay,N. and Santella,R.M. (2010) Aberrant promoter hypermethylation and genomic hypomethylation in tumor adjacent normal tissues and blood from breast cancer patients. *Anticancer Res.*, **30**, 2489–2496.
- Avvedimento,E.V., Obici,S., Sanchez,M., Gallo,A., Musti,A. and Gottesman,M.E. (1989) Reactivation of thyroglobulin gene expression in transformed thyroid cells by 5-azacytidine. *Cell*, **58**, 1135–1142.
- Angrisano,T., Sacchetti,S., Natale,F., Cerrato,A., Pero,R., Keller,S., Peluso,S., Perillo,B., Avvedimento,V.E., Fusco,A. *et al.* (2011) Chromatin and DNA methylation dynamics during retinoic acid-induced RET gene transcriptional activation in neuroblastoma cells. *Nucleic Acids Res.*, **39**, 1993–2006.
- Halazonetis,T.D., Gorgoulis,V.G. and Bartek,J. (2008) An oncogene-induced DNA damage model for cancer development. *Science*, **319**, 1352–1355.
- Lengauer,C., Kinzler,K.W. and Vogelstein,B. (1997) DNA methylation and genetic instability in colorectal cancer cells. *Proc. Natl Acad. Sci. USA*, **94**, 2545–2550.
- Cuozzo,C., Porcellini,A., Angriano,T., Morano,A., Lee,B., Di Pardo,A., Messina,S., Iuliano,R., Fusco,A., Santillo,M.R. *et al.* (2007) DNA damage homology-directed repair and DNA methylation. *PLoS Genet.*, **3**, e110.
- Jasin,M. (1996) Genetic manipulation of genomes with rare-cutting endonucleases. *Trends Genet.*, **12**, 224–228.
- Pierce,A.J., Johnson,R.D., Thompson,L.H. and Jasin,M. (1999) XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.*, **13**, 2633–2638.
- Rountree,M.R. and Selker,E.U. (1997) DNA methylation inhibits elongation but not initiation of transcription in *Neurospora crassa*. *Genes Dev.*, **11**, 2383–2395.
- O'Hagan,H.M., Mohammad,H.P. and Baylin,S.B. (2008) Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet.*, **4**, e1000155.
- Lee,G.E., Kim,J.H., Taylor,M. and Muller,M.T. (2010) DNA methyltransferase 1-associated protein (DMAP1) is a co-repressor that stimulates DNA methylation globally and locally at sites of double strand break repair. *J. Biol. Chem.*, **285**, 37630–37640.

15. Ha, K., Lee, G.E., Pali, S.S., Brown, K.D., Takeda, Y., Liu, K., Bhalla, K.N. and Robertson, K.D. (2011) Rapid and transient recruitment of DNMT1 to DNA double-strand breaks is mediated by its interaction with multiple components of the DNA damage response machinery. *Hum. Mol. Genet.*, **20**, 126–140.
16. Sandovici, I., Kassovska-Bratinova, S., Vaughan, J.E., Stewart, R., Leppert, M. and Sapienza, C. (2006) Human imprinted chromosomal regions are historical hot-spots of recombination. *PLoS Genet.*, **2**, e101.
17. Sigurdsson, M.I., Smith, A.V., Bjornsson, H.T. and Jonsson, J.J. (2009) HapMap methylation-associated SNPs markers of germline DNA methylation positively correlate with regional levels of human meiotic recombination. *Genome Res.*, **19**, 581–589.
18. Lee, B., Morano, A., Porcellini, A. and Muller, M.T. (2012) GADD45 $\alpha$  inhibition of DNMT1 dependent DNA methylation during homology directed DNA repair. *Nucleic Acids Res.*, **40**, 2481–2493.
19. Gong, X.Q., Nedialkov, Y.A. and Burton, Z.F. (2004) Alpha-amanitin blocks translocation by human RNA polymerase II. *J. Biol. Chem.*, **279**, 27422–27427.
20. Jones, E., Kimura, H., Vigneron, M., Wang, Z., Roeder, R.G. and Cook, P.R. (2000) Isolation and characterization of monoclonal antibodies directed against subunits of human RNA polymerases I, II, and III. *Exp. Cell Res.*, **254**, 163–172.
21. Ioannidis, P., Havredaki, M., Courtis, N. and Trangas, T. (1996) *In vivo* generation of 3' and 5' truncated species in the process of c-myc mRNA decay. *Nucleic Acids Res.*, **24**, 4969–4977.
22. Leclerc, G.J., Leclerc, G.M. and Barredo, J.C. (2002) Real-time RT-PCR analysis of mRNA decay: half-life of Beta-actin mRNA in human leukemia CCRF-CEM and Nalm-6 cell lines. *Cancer Cell Int.*, **2**, 1.
23. Walton, E.L., Francastel, C. and Velasco, G. (2011) Maintenance of DNA methylation: Dnmt3b joins the dance. *Epigenetics*, **6**, 1373–1377.
24. Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K. *et al.* (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature*, **450**, 908–912.
25. Meilinger, D., Fellinger, K., Bultmann, S., Rothbauer, U., Bonapace, I.M., Klinkert, W.E., Spada, F. and Leonhardt, H. (2009) Np95 interacts with de novo DNA methyltransferases, Dnmt3a and Dnmt3b, and mediates epigenetic silencing of the viral CMV promoter in embryonic stem cells. *EMBO Rep.*, **10**, 1259–1264.
26. Rottach, A., Frauer, C., Pichler, G., Bonapace, I.M., Spada, F. and Leonhardt, H. (2010) The multi-domain protein Np95 connects DNA methylation and histone modification. *Nucleic Acids Res.*, **38**, 1796–1804.
27. Papait, R., Pistore, C., Negri, D., Pecoraro, D., Cantarini, L. and Bonapace, I.M. (2007) Np95 is implicated in pericentromeric heterochromatin replication and in major satellite silencing. *Mol. Biol. Cell*, **18**, 1098–1106.
28. Hervouet, E., Lallier, L., Debie, E., Cheray, M., Geairon, A., Rogniaux, H., Loussouarn, D., Martin, S.A., Vallette, F.M. and Cartron, P.F. (2010) Disruption of Dnmt1/PCNA/UHRF1 interactions promotes tumorigenesis from human and mice glial cells. *PLoS One*, **5**, e11333.
29. Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F. *et al.* (2007) Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature*, **445**, 671–675.
30. Jin, S.G., Guo, C. and Pfeifer, G.P. (2008) GADD45A does not promote DNA demethylation. *PLoS Genet.*, **4**, e1000013.
31. Helleday, T., Lo, J., van Gent, D.C. and Engelward, B.P. (2007) DNA double-strand break repair: from mechanistic understanding to cancer treatment. *DNA Repair (Amst)*, **6**, 923–935.
32. Métivier, R., Gallais, R., Tiffocche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G. *et al.* (2008) Cyclical DNA methylation of a transcriptionally active promoter. *Nature*, **452**, 45–50.
33. Mutsaers, V. and Felsenfeld, G. (2004) Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. *EMBO J.*, **23**, 138–149.
34. Hinshelwood, R.A., Melki, J.R., Huschtscha, L.I., Paul, C., Song, J.Z., Stirzaker, C., Reddel, R.R. and Clark, S.J. (2009) Aberrant de novo methylation of the p16INK4A CpG island is initiated post gene silencing in association with chromatin remodelling and mimics nucleosome positioning. *Hum. Mol. Genet.*, **18**, 3098–3109.
35. Stirzaker, C., Song, J.Z., Davidson, B. and Clark, S.J. (2004) Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res.*, **64**, 3871–3877.
36. Kangaspeka, S., Stride, B., Métivier, R., Polycarpou-Schwarz, M., Ibberson, D., Carmouche, R.P., Benes, V., Gannon, F. and Reid, G. (2008) Transient cyclical methylation of promoter DNA. *Nature*, **452**, 112–115.
37. Perillo, B., Ombra, M.N., Bertoni, A., Cuozzo, C., Sacchetti, S., Sasso, A., Chiariotti, L., Malorni, A., Abbondanza, C. and Avvedimento, E.V. (2008) DNA oxidation as triggered by H3K9me2 demethylation drives estrogen-induced gene expression. *Science*, **319**, 202–206.
38. Le May, N., Fradin, D., Iltis, I., Bougnères, P. and Egly, J.M. (2012) XPG and XPF endonucleases trigger chromatin looping and DNA demethylation for accurate expression of activated genes. *Mol. Cell*, **47**, 622–632.
39. Chen, Z.X. and Riggs, A.D. (2011) DNA methylation and demethylation in mammals. *J. Biol. Chem.*, **286**, 18347–18353.
40. Sen, G.L., Reuter, J.A., Webster, D.E., Zhu, L. and Khavari, P.A. (2010) DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature*, **463**, 563–567.
41. Radulescu, I., Elmroth, K. and Stenérlov, B. (2004) Chromatin organization contributes to non-randomly distributed double-strand breaks after exposure to high-LET radiation. *Radiat. Res.*, **16**, 1–8.
42. Citterio, E., Papait, R., Nicassio, F., Vecchi, M., Gomiero, P., Mantovani, R., Di Fiore, P.P. and Bonapace, I.M. (2004) Np95 is a histone-binding protein endowed with ubiquitin ligase activity. *Mol. Cell Biol.*, **24**, 2526–2535.